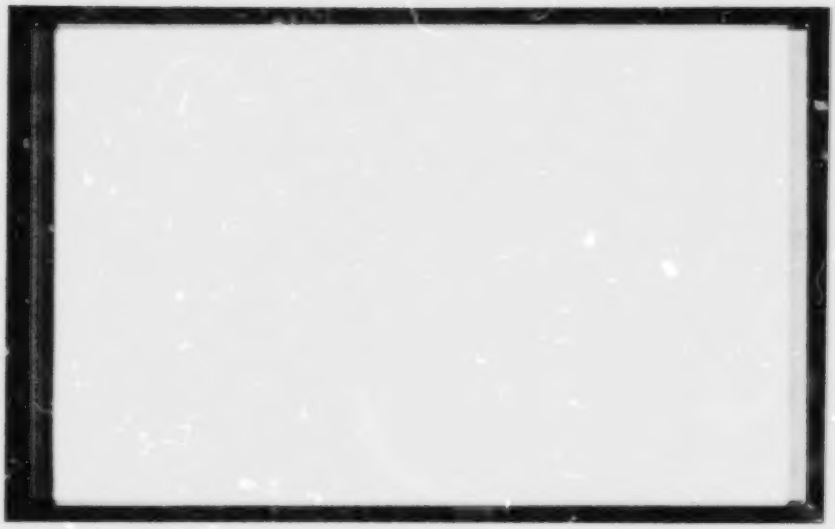
UNIVERSITÉ DE MONTRÉAL



CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

DEUX PROCÉDURES D'ESTIMATION EN PRÉSENCE D'AUTOCORRÉLATION SPATIALE DANS LES RÉSIDUS

par

Denis Bolduc

Thèse de doctorat, département de sciences économiques,
dirigée par Marcel G. Dagenais

Université de Montréal
Centre de recherche sur les transports - Publication #443
Juin 1985

Dépôt légal - 1er trimestre 1986

TABLE DES MATIERES

SOMMAIRE xi

INTRODUCTION 1

CHAPITRE I. Les processus stochastiques spatiaux. 5

1) Concepts et définitions. 7

1.1) Généralités. 7

Définition 1. Processus stochastique discret défini sur $\mathbb{Z}^2$ (champ aléatoire) 7

Définition 2. Champ aléatoire stationnaire au sens strict. . . 8

Définition 3. Champ aléatoire stationnaire au sens large (S.S.L.) 9

1.2) Concepts de séries temporelles et d'analyse spectrale. 9

Définition 4. Convergence en moyenne quadratique 10

a) Processus autorégressif d'ordre p AR(p) 12

b) Processus moyenne mobile d'ordre q MA(q). 16

1.3) Concepts d'analyse spectrale 18

a) Spectre d'un bruit blanc 20

b) Une propriété du spectre 20

c) Spectre d'un MA(1) 21

2) Modèles autorégressifs spatiaux. 22

2.1) Deux modèles autorégressifs spatiaux stationnaires 22

2.2) Modèle spatial moyenne mobile et concept d'isotropie 29

2.3) La spécification dans les modèles autorégressifs spatiaux. . . . 33

2.3.1) Le corrélogramme spatial empirique 34

2.3.2) La spécification dans les deux modèles autorégressifs isotropes de Haining. 38

3) Modèles autorégressifs spatiaux pour une configuration irrégulière 42

3.1) Concepts et notation. 43

3.2) Les matrices de contiguïté normalisées. 49

CHAPITRE II. Tests d'autocorrélation spatiale. 53

1) Mesures d'autocorrélation spatiale pour des données continues . . . 55

1.1) Mesure de Moran 56

1.2) Mesure de Geary 57

1.3) Invariance topologique. 57

1.4) Mesure de Cliff et Ord. 58

1.5) Formulation matricielle de I_x^M et I_x^{CO}. 59

1.6) Théorie de distribution de I_c^{CO} et I_x^M. 60

1.6.1) Moments de la loi de distribution d'échantillonnage de I_x^{CO} 60

1.6.1.1) Preuve de l'applicabilité du théorème Pitman-Koopmans 61

1.6.1.2) Dérivation des deux premiers moments de distribution de I_x^{CO} 62

1.6.2) Equivalence entre (22) et (1.38) de Cliff et Ord (1981). 64

1.6.3) Théorie de distribution asymptotique de I_x^{CO} 66

1.6.3.1) Digression statistique. 67

A.1) Fonction génératrice de moments (f.g.m.) 67

A.2) Appliqué à la normale. 69

B.1) Fonction génératrice de cumulants (f.g.c.) . . . 70

B.2) Appliqué à la normale centrée (non réduite). . . 71

1.6.3.2) Normalité asymptotique de I_x^{CO} 72

1.6.4) Généralisation du test I_x^{CO} pour un ordre g. 80

2) L'autocorrélation spatiale dans les résidus 82

2.1) Les résidus MCO 82

2.1.1) Normalité asymptotique. 84

2.2) Procédures alternatives 85
La classe LUS. 85
- Linéarité 85
- Non biais 86
- Matrice v.c. scalaire 86
Résidus BLUS 87
2.3) Expériences de Monté-Carlo. 89
2.3.1) Le ratio des vraisemblances 90

CHAPITRE III. Estimation du SAR(1) résiduel. 91

1) Quelques schémas de dépendance spatiale dans les résidus. 94
1.1) Le modèle spatial bilatéral unidimensionnel : cas ① 94
1.2) Le modèle autorégressif quadrilataire d'ordre 1 : QA(1) : cas ② 95
1.3) Le modèle autorégressif spatial en vague : cas ③ 98
1.4) Configuration spatiale en étoile : cas ④ 100
1.5) Contiguïté spatiale totale : cas ⑤ 102
1.6) Contiguïté spatiale standard : cas ⑥ 103
2) Estimation de ρ par la méthode des moindres carrés ordinaires (MCO) . 104
3) Méthode itérative Cochrane-Orcutt (Ord (1975)). 108
3.1) Estimation $\tilde{\rho}$ du maximum de vraisemblance. 109
4) Méthode du maximum de vraisemblance (Hepple (1976)) 112
5) Etude du comportement analytique de la fonction de vraisemblance du SAR(1) 115
5.1) Comportement analytique de L** pour une matrice de contiguïté W quelconque 116
5.2) Comportement analytique de L** pour une matrice de contiguïté M. 124
5.2.1) Racine caractéristique dominante de la matrice de contiguïté M 124
5.2.2) Comportement de la fonction L** dans la région stable. 126
5.3) Comportement analytique de L** pour une matrice de contiguïté W. 130
6) Quelques résultats empiriques concernant l'estimation du SAR(1) résiduel. 132
6.1) Méthodes d'estimation inspirées de la série chronologique . . . 133
6.2) Résultats d'estimation du maximum de vraisemblance pour le SAR(1) résiduel. 135

7) Sur les propriétés asymptotiques des estimateurs du maximum de vraisemblance pour un modèle SAR(1) résiduel 137

7.1) Résultats de Magnus (1978) concernant les propriétés des estimateurs (MCG) du maximum de vraisemblance 138

7.2) Résultats de Magnus (1978) pour un modèle SAR(1) résiduel . . 140

Résultat 1 141

Résultat 2 143

CHAPITRE IV. Estimation d'un processus AR(2) résiduel spatial autorégressif 145

1) Fonction de vraisemblance 146

2) Dérivées premières. 147

3) Dérivées secondes 149

4) Evaluation de la matrice variance-covariance. 151

5) Formules programmées pour les dérivées secondes. 152

6) Modification de l'algorithme suite à l'imposition de la normalisation symétrique. 153

6.1) Dérivées premières. 154

6.2) Dérivées secondes 155

CHAPITRE V. Une application : l'accessibilité et la structure spatiale de l'économie de la république d'Irlande. 159

Résultats de la régression. 162

CHAPITRE VI. Etude et implication de différentes formes fonctionnelles binaires pour les w_{ij} 169

1) La forme binaire (0, 1) standard (non normalisée) 171

2) La forme binaire (0, 1) normalisée. 178

2.1) Résultats de l'estimation pour une matrice de contiguïté booléenne normalisée. 179

3) La forme binaire (0, 1) avec normalisation symétrique. 181

3.1) Résultats de l'estimation pour une matrice de contiguïté booléenne avec normalisation symétrique 183

CHAPITRE VII. Estimation des paramètres d'une forme fonctionnelle w_{ij}. . 187

1) Facteurs qui peuvent expliquer la structure spatiale des résidus. . 188

1.1) Autorégression résiduelle 188

1.2) La distance 189

1.3) Autres facteurs 189

2) $w_{ij} = d_{ij}^{-\theta_1}$ 190

2.1) La forme non normalisée avec $w_{ij} = d_{ij}^{-\theta_1}$ 190

2.2) $w_{ij} = d_{ij}^{-\theta_1}$ avec normalisation symétrique 193

2.2.1) Allure de la fonction de vraisemblance 193

2.2.2) Interprétation économique. 196

2.2.3) Résultats de l'estimation. 199

2.3) $w_{ij} \neq 0$ si i et j sont contigües 201

2.3.1) Estimation du modèle $w_{ij} = d_{ij}^{-\theta_1}$, si i et j sont contigües avec normalisation symétrique 203

3) Forme généralisée $w_{ij} = d_{ij}^{-\theta_1} \cdot p_{ij}^{\theta_2}$ 204

3.1) Interprétation économique. 206

3.2) Résultats d'estimation pour

$$w_{ij} = d_{ij}^{-\theta_1} \cdot \left(\frac{1}{|\ln Rev_1 - \ln Rev_j|}\right)^{\theta_2}$$ 207

CHAPITRE VIII. Formulation autorégressive pour des données de flux. . . 211

Formulation du schéma autorégressif 212

CHAPITRE IX. Algorithme d'estimation d'une forme autorégressive résiduelle pour des données de flux 217

1) Fonction de vraisemblance. 218

2) Dérivées premières 219

3) Dérivées secondes. 220

4) Formules à programmer pour les dérivées secondes 222

CHAPITRE X. Les moindres carrés généralisés (MCG) pour un problème hétérospatial 229

1) Décomposition en deux étapes. 230

2) Quelques propriétés des estimateurs moindres carrés généralisés . . 234

2.1) Ω connu 234

2.2) Ω inconnu 236

2.3) Propriétés de l'estimateur des moindres carrés ordinaires. . 237

3) L'estimateur $\tilde{\beta}$ du maximum de vraisemblance. 238

4) Procédure d'estimation en deux étapes 240

4.1) Modélisation convergente de l'hétéroscédasticité. 240

4.1.1) La forme hétéroscédastique multiplicative 241

4.1.2) Estimation de la forme hétéroscédastique multiplicative. 245

4.2) Modélisation de la corrélation spatiale 246

4.2.1) Une procédure MCGE. 247

4.2.2) Une forme fonctionnelle raisonnable pour la modélisation de la corrélation spatiale 249

4.2.2.1) Quelques variables pour expliquer la corrélation spatiale 253

i) Des variables à doubles indices 253

ii) Autres variables. 254

4.2.3) Une procédure du maximum de vraisemblance 255

5) Sur les propriétés asymptotiques des estimateurs du maximum de vraisemblance pour le modèle MCG spatial 258

CHAPITRE XI. Méthode des moindres carrés généralisés (MCG) du maximum de vraisemblance comme solution à un problème de corrélation spatiale 259

1) Fonction de vraisemblance et dérivées premières 261

2) Dérivées secondes 263

3) Simplification des formules pour la programmation 265

4) Calcul de la matrice variance-covariance des paramètres estimés . . 265

5) Changement d'espace pour les paramètres 267

CHAPITRE XII. Résultats d'estimation de la procédure des moindres carrés généralisés du maximum de vraisemblance 269

1) Résultats d'estimation de la forme $r_{ij} = \tanh (\alpha_1 d_{ij}^{-\lambda})$. 270

2) Autres spécifications

2.1) $Z'_{ij} \alpha = d_{ij}^{-\lambda} (\alpha_1 + \alpha_2 V_{ij}^1)$ 278

2.1.1) $V_{ij}^1 = 1$, si $Pop_i > \overline{Pop}$ et $Pop_j > \overline{Pop}$; = 0 sinon . . 278

2.1.2) $V_{ij}^1 = 1$, si $Pop_i > \overline{Pop}$ ou $Pop_j > \overline{Pop}$; = 0 sinon . . 282

2.1.3) $V_{ij}^1 = Pop_i \cdot Pop_j$ 283

2.1.4) $V_{ij}^1 = 1$, si i et j sont contiguës ; = 0 sinon . . . 283

2.1.5) $V_{ij}^1 = \frac{1}{|\ln Rev_i - \ln Rev_j|}$ 284

3) Evaluation générale de nos estimations. 286

CONCLUSION 287

APPENDICE I. Données 291

APPENDICE II. Résultats concernant les statistiques d'autocorrélation spatiale 297

1) Théorème Pitman-Koopmans 298

1.1) Statistique exhaustive totale et statistique libre 299

1.1.1) Statistique exhaustive totale. 299

Définition 1 : Statistique exhaustive (sufficient statistic) 299

Définition 2 : Modèle total et statistique totale 301

1.1.2) Statistique libre. 302

Définition 3 : Statistique libre. 302

1.2) Démonstration du théorème Pitman-Koopmans. 302

2) Espérance mathématique d'une forme quadratique 304

3) Dérivation des moments de distribution des statistiques d'autocorrélation spatiale sur les résidus 305

APPENDICE III. Concepts utiles lors de l'étude des propriétés asymptotiques des estimateurs. 309

1) Rappel sur certaines propriétés d'échantillon fini 310

2) Théorie asymptotique 312

2.1) La limite en probabilité 312

3) Loi de distribution limite 314

3.1) Convergence en distribution. 315

3.2) Propositions concernant l'établissement de la loi de distribution limite. 316

4) Application dans le cas de l'estimation MCO. 317

5) Efficacité asymptotique. 319

APPENDICE IV. Propriétés des estimateurs du maximum de vraisemblance . . 321

1) Echantillon de taille finie. 322

2) Echantillon asymptotique 322

REMERCIEMENTS 324

BIBLIOGRAPHIE 325

SOMMAIRE

Dans le but de considérer le problème de corrélation dans les résidus de modèles de régression avec données spatiales (coupe transversale), on suggère ici deux approches basées sur le principe des moindres carrés généralisés. Selon la première approche, on généralise la formulation autorégressive spatiale d'ordre (1) [SAR(1)] développée par les auteurs Cliff et Ord (1975c). Afin de prendre en considération la structure physique de contiguïté et le caractère économico-géographique des N zones de l'étude, cette formulation autorégressive exprime l'erreur ε_i comme une somme pondérée sur certaines autres erreurs de la régression. Matriciellement, on a : $y = X\beta + \varepsilon$ avec $\varepsilon = \rho W \varepsilon + u$, où $u \sim (0, \sigma^2 I_N)$, ρ est un coefficient d'autocorrélation spatiale et W une matrice (N×N) de poids qui permet de trouver le délai autorégressif spatial d'ordre 1. Les w_{ij} sont fonctions d'un vecteur de paramètre dont les valeurs ont jusqu'à maintenant toujours été fixées à priori. Ces formes fonctionnelles peuvent s'écrire : $w_{ij} = g\,(V_{ij}^1, V_{ij}^2, \ldots, V_{ij}^P, \theta_1, \theta_2, \ldots, \theta_H)$, c'est-à-dire que ce sont des termes qui dépendent de H paramètres et de P variables V_{ij}^P pertinentes.

Soit λ_{MAX} la racine caractéristique maximale de la matrice W, nous montrons que le modèle SAR(1) est stable[1] lorsque $|\rho| < 1/\lambda_{MAX}$. Nous proposons au chapitre III deux types de normalisation sur la matrice W tel que le modèle SAR(1) est stable lorsque $|\rho| < 1$. En exploitant une de ces

---

[1]. Au chapitre I, nous montrons que pour des configurations irrégulières, contrairement à la pratique, on doit parler de stabilité spatiale plutôt que de stationnarité spatiale.

deux normalisations, comme généralisation du SAR(1), nous proposons au chapitre IV une procédure du maximum de vraisemblance qui vise l'estimation jointe des paramètres des fonctions w_{ij} et de ceux de la régression. Les résultats empiriques des chapitres V à VII et les résultats théoriques du chapitre III, nous portent à conclure que cette normalisation est des deux points de vue précédents, désirable. Sur le plan empirique sur le domaine stable $|\rho| < 1$, cette normalisation en rendant lisse la fonction de vraisemblance facilite la tâche de l'algorithme dans la recherche d'une solution optimale. Sur le plan théorique en plus de donner à ρ une interprétation économique intéressante $[-1 < \rho < 1]$, cette normalisation, lorsque ρ_o (la valeur vraie de ρ) appartient au domaine unitaire, assure la propriété de convergence aux estimations du maximum de vraisemblance.

Aux chapitres X à XII, nous développons une procédure du maximum de vraisemblance selon laquelle nous nous intéressons à estimer les éléments de la matrice des corrélations des résidus. Après avoir corrigé, si nécessaire, l'hétéroscédasticité, le modèle homoscédastique transformé s'écrit $y^* = X^*\beta + \varepsilon^*$ avec $E(\varepsilon^*) = 0$ et $E(\varepsilon^*\varepsilon^{*\prime}) = \sigma^2 R$, R étant la matrice $(N \times N)$ des corrélations entre les résidus de l'étude. Afin de modéliser les corrélations spatiales r_{ij}, nous postulons la forme suivante : $r_{ij} = \tanh(z'_{ij}\,\alpha)$ avec $z'_{ij}\,\alpha = d_{ij}^{-\lambda}(\alpha_1 + \alpha_2 V_{ij}^2 + \ldots + \alpha_H V_{ij}^H)$, d_{ij} représentant la distance entre i et j. Grâce à cette forme, r_{ij} est limité à l'intervalle unitaire $[-1 \leq r_{ij} \leq 1]$. De cette façon, nous imposons de plus la propriété désirable de corrélation nulle entre les erreurs de zones très éloignées. Le problème principal de notre approche réside dans le fait qu'on ne peut garantir

pour la matrice R la propriété de matrice définie positive. Les résultats d'estimation de cette procédure sont pour leur part assez décevants. La spécification la plus générale que l'on a pu estimer de façon satisfaisante s'est limitée à $z'_{ij}\ \alpha = c_1\ d_{ij}^{-\lambda}$.

INTRODUCTION

Dans la formulation de modèles économétriques qui visent l'étude d'un problème de coupe transversale (données spatiales), les économètres négligent habituellement le fait que les erreurs de régression sont fort probablement correlées entre elles. Par exemple, dans un modèle keynesien d'explication de la consommation d'une zone i par le revenu disponible de cette même zone, on ignore, à tort, le fait que le niveau de consommation de cette zone i dépend aussi très certainement du revenu des zones avoisinantes. Si ce phénomène n'est pas explicitement introduit dans l'ensemble que forme les régresseurs, il se réflètera alors dans la structure des résidus de la régression (autocorrélation spatiale dans les résidus). Si le problème est présent et qu'il est ignoré par le chercheur, l'estimation par la procédure des moindres carrés ordinaires, bien que non biaisée, est inefficace. Comme alternative, il faut postuler une formulation de type MCG (moindres carrés généralisés).

Ainsi que le titre le souligne, nous allons proposer deux procédures MCG d'estimation en présence d'autocorrélation spatiale résiduelle. La première est issue d'une formulation autorégressive spatiale inspirée des modèles ARMA temporels. Dans la seconde, par le biais d'estimation de formes fonctionnelles, nous chercherons à modéliser directement les corrélations entre les résidus de la régression. D'après les résultats de notre recherche, la procédure autorégressive spatiale risque d'intéresser grandement les économètres. Pour la seconde procédure, l'intérêt principal sera probablement d'avoir ouvert une voie pour d'autres recherches.

La première procédure étant basée sur une formulation de type autoregressif, dans le but de donner à notre approche un fondement théorique adéquat, nous étudions au chapitre I la théorie des champs aléatoires (les processus stochastiques en deux dimensions). L'intérêt de ce chapitre est double. D'une part, on veut présenter les principaux concepts méthodologiques qui ont été développés afin de permettre l'étude de tels processus. Ceci nous permettra alors de constater la présence de deux tendances. D'autre part, en ce qui concerne le modèle qui nous intéressera dans les chapitres subséquents, cet effort contribuera à le situer de façon bien précise, par rapport à ces deux courants méthodologiques. Du même coup, on pourra ainsi faire ressortir clairement les hypothèses sous-jacentes de chacune des ces approches.

Afin d'évaluer de façon statistique l'importance ou la présence du problème d'autocorrélation spatiale, nous présentons au chapitre II des résultats qui concernent les tests d'autocorrélation spatiale. De façon générale, pour exécuter ces tests à justification asymptotique, il suffit de comparer la statistique calculée à une $N(0, 1)$ à $(1-\alpha)$ % .

Au chapitre III, on étudiera les propriétés théoriques et empiriques liées à une formulation autorégressive spatiale d'ordre 1 qui, lorsqu'appliquée aux N résidus d'une régression, s'exprime comme suit :

$$y = X\beta + \varepsilon$$

$$\varepsilon = \rho W \varepsilon + u$$

où ρ est un coefficient d'autocorrélation et W une matrice (N×N) qui permet d'introduire de l'information sur le caractère économico-géographique de la région étudiée. Jusqu'à maintenant, les w_{ij} ont toujours été considérés comme des formes fonctionnelles fixées a priori. Après avoir vanté les avantages de certaines normalisations sur la matrice W, comme généralisation, nous proposerons au chapitre IV une procédure du maximum de vraisemblance qui vise l'estimation jointe des paramètres des w_{ij} et de ceux de la régression. Aux chapitres V à VII, nous appliquerons notre approche au modèle de régression qui nous intéresse.

De façon à présenter l'objet de nos futures recherches, nous développerons aux chapitres VIII et IX l'extension de cette approche pour des données de flux entre les zones i et j de l'étude.

Nous consacrerons les chapitres X à XII à l'élaboration de la seconde procédure. Comme souligné, selon cette méthodologie, nous chercherons à modéliser des corrélations résiduelles.

Au chapitre X, nous proposerons des formes fonctionnelles qui risquent d'être pertinentes dans la modélisation des corrélations entre les résidus de la régression. Au chapitre XI, nous développerons pour ce cas, une procédure du maximum de vraisemblance. C'est au chapitre XII que nous procéderons à l'étude des estimations empiriques issues de cette formulation.

CHAPITRE I

LES PROCESSUS STOCHASTIQUES SPATIAUX

Les nombreuses années de recherche visant l'applicabilité de la théorie des processus stochastiques en une dimension à l'étude des séries temporelles ont finalement débouché sur l'implantation d'une procédure méthodologique qui, on le sait, est devenue extrêmement populaire. Ce même courant de recherche a poussé des auteurs comme Whittle (1954) à développer la théorie des processus stochastiques multidimensionnels. Lorsque limitée à 2 dimensions, cette théorie s'appelle aussi la théorie des champs aléatoires.

Selon cette dernière, on s'intéresse à la loi de probabilité d'une séquence stochastique formée par des variables aléatoires à doubles indices $X_{i,j}$. On pourra s'apercevoir que le domaine des données spatiales (coupe transversale) constitue la sphère d'application idéale pour cette théorie, d'où le nom : processus stochastiques spatiaux.

L'intérêt de ce chapitre est double. D'une part, on veut présenter les principaux concepts méthodologiques qui ont été développés afin de permettre l'étude de tels processus. Ceci nous permettra alors de constater la présence de deux tendances. D'autre part, en ce qui concerne le modèle qui nous intéressera dans les chapitres subséquents, cet effort contribuera à le situer de façon bien précise, par rapport à ces deux courants méthodologiques. Du même coup, on pourra ainsi faire ressortir clairement les hypothèses sous-jacentes de chacune de ces approches.

A la première section, on définira formellement ce qu'est un processus stochastique spatial discret[1]. Dans un objectif purement pédagogique, on rap-

---

[1]. En économie, les processus sont généralement de type discrets, c'est-à-dire qu'on considère la réalisation de variables aléatoires pour un nombre fini d'emplacements. Dans le cas continu, on parlerait plutôt de processus spatiaux continus.

pellera alors certains concepts qui touchent l'étude des séries chronologiques (temporelles). A la section 2, on présentera l'extension en deux dimensions de ces derniers concepts. A la section finale, on se concentrera sur les outils méthodologiques de la seconde tendance, pour, par la suite, passer au développement théorique du modèle que l'on veut estimer.

1) Concepts et définitions[1]

1.1) Généralités

Soit $\mathbb{Z}^2$ le plan des entiers. On a :

Définition 1. *Processus stochastique discret défini sur $\mathbb{Z}^2$ (champ aléatoire)*

C'est la collection telle qu'à chaque élément i,j de $\mathbb{Z}^2$, est associée une variable aléatoire $X_{i,j}$, ce que l'on note :

$$\{X_{i,j} : i,j \in \mathbb{Z}^2\} \quad . \tag{1}$$

Dans la définition d'un processus stochastique, on considère qu'il existe un mécanisme qui génère les $X_{i,j}$. Un des buts principaux de l'analyste consiste à identifier ce mécanisme.

Afin de caractériser la loi de probabilité du processus (1), on considère l'étude des lois de tout sous-ensemble fini $X_{i_1,j_1}, \ldots, X_{i_{PQ},j_{PQ}}$ $P \in \mathbb{N}$, $Q \in \mathbb{N}$, $\{i_1,j_1, i_2,j_2,\ldots,i_{PQ},j_{PQ}\} \in \mathbb{Z}^2$ qui se réalise sur une sous-région

---

[1]. Voir Granger (1969) ou Haining (1980).

finie Γ [1]. La fonction de distribution jointe de $\{X_{i_1,j_1},\ldots,X_{i_{PQ},j_{PQ}}\}$ s'écrit :

$$(2)\quad F(x_{i_1,j_1},\ldots,x_{i_{PQ},j_{PQ}}) = P(X_{i_1,j_1} < x_{i_1,j_1},\ldots,X_{i_{PQ},j_{PQ}} < x_{i_{PQ},j_{PQ}}) \ .$$

Comme, en pratique, on devra inférer les propriétés du processus (1) à partir d'un seul échantillon observé $[x_{i_1,j_1},\ldots,x_{i_{PQ},j_{PQ}}]$ [2], on se doit, tout comme en séries chronologiques, d'imposer la stationnarité sur le processus (1).

Définition 2. *Champ aléatoire stationnaire au sens strict*

Un champ aléatoire $\{X_{i,j} : i,j \in \mathbb{Z}^2\}$ est stationnaire au sens strict (S.S.S.) si :

$$(3)\quad F(x_{i_1+s,j_1+t},\ldots,x_{i_{PQ}+s,j_{PQ}+t}) = F(x_{i_1,j_1},\ldots,x_{i_{PQ},j_{PQ}}) \ ,$$

$$\forall PQ, \quad \forall s \text{ et } t, \text{ et } \forall \{i_1,j_1,\ldots,i_{PQ},j_{PQ}\} \ .$$

Ainsi, la loi de probabilité jointe du vecteur $X_{(PQ\times 1)} = [X_{i_1,j_1},\ldots,X_{i_{PQ},j_{PQ}}]'$ dont la réalisation se produit sur le plan rectangulaire Γ est la même , quel que soit l'emplacement de Γ dans le plan $\mathbb{Z}^2$. C'est grâce à cette invariance par rapport à la localisation, que les caractéristiques de la loi de distribution jointe du processus $\{X_{i,j} : i,j \in \mathbb{Z}^2\}$

---

[1]. Etant sur un plan rectangulaire, tout sous-ensemble fini de (1) forme une sous-région Γ de P emplacements sur le plan vertical et Q sur le plan horizontal. Par convention, on note P X Q = PQ.

[2]. Le vecteur $x_{(PQ\times 1)}$ est bien entendu la réalisation du vecteur aléatoire $X_{(PQ\times 1)}$: l'échantillon observable.

peuvent être décrites complètement. Dans la pratique, on se borne habituellement à ne caractériser que les deux premiers moments de la loi de distribution du processus. Ceci nous incite donc à définir la stationnarité au sens large.

Définition 3. *Champ aléatoire stationnaire au sens large (S.S.L.)*

Un champ aléatoire $\{X_{i,j} : i,j \in \mathbb{Z}^2\}$ est S.S.L. si :

i) $E(X_{i,j}) = \mu$, $\forall\ i,j \in \mathbb{Z}^2$,

ii) $V(X_{i,j}) = \sigma(0,0) < \infty$, $\forall\ i,j \in \mathbb{Z}^2$,

iii) $COV(X_{i,j}, X_{i+s,j+t}) = \sigma(s,t)$, $\forall\ i,j \in \mathbb{Z}^2$,

avec $\sigma(s,t) = \sigma(-s,-t)$, où on a comme corrélations :

iv) $\rho(s,t) = \dfrac{\sigma(s,t)}{\sigma(0,0)} = \rho(-s,-t)$, $s,t = 0, \pm1, \pm2, \ldots$

c'est-à-dire que les corrélations ne dépendent que des délais s et t et non de la localisation i,j. La série $\{\rho(s,t)\}$ est appelée série d'autocorrélation du cham $\{X_{i,j} : i,j \in \mathbb{Z}^2\}$ ou corrélogramme du champ $\{X_{i,j} : i,j \in \mathbb{Z}^2\}$.

On peut évidemment vérifier que la stationnarité au sens strict, définie à l'équation (3), implique aussi les points i) à iv) précédents.

1.2) Concepts de séries temporelles et d'analyse spectrale

Afin de simplifier l'extension des concepts de processus stochastiques en une dimension à ceux des champs aléatoires, on veut résumer ici certains instruments qui sont utilisés en analyse des séries temporelles (domaine du temps) et en analyse spectrale (domaine des fréquences)[1].

---

[1]. Pour une présentation plus détaillée, consulter le chapitre V de Gourieroux et Monfort (1983).

Un processus stochastique discret $\{X_t : t \in \mathbb{Z}\}$ est stationnaire au sens large si : $E(X_t) = \mu \ \forall t$, $V(X_t) = \sigma_x^2 < \infty \ \forall t$ et $\text{cov}(X_t, X_{t-\tau}) = \lambda_\tau = \lambda_{-\tau}$. Dans le but d'alléger la présentation, on fait l'hypothèse que $E(X_t) = 0$. D'après notre notation, $V(X_t) = \sigma_x^2 = \lambda_0$. Dans les pages qui suivent, nous aurons grandement besoin de :

Définition 4. *Convergence en moyenne quadratique*[1]

Une suite X_N de variables, vérifiant $E(X_N^2) < \infty$, converge en moyenne quadratique vers une variable X, si $\lim_{N\to\infty} E[(X_N - X)^2] = 0$. Alors $\lim_{N\to\infty} E(X_N^2) = E(X^2)$, où $E(X_N^2) < \infty \ \forall N$.

Proposition 1

Si $\{X_t : t \in \mathbb{Z}\}$ est un processus stationnaire et si $\{a_i : i \in \mathbb{Z}\}$ est une suite de nombres réels absolument convergente : $\sum_{i=-\infty}^{\infty} |a_i| < +\infty$, les égalités $Y_t = \sum_{i=-\infty}^{\infty} a_i X_{t-i}$, $t \in \mathbb{Z}$ définissent un nouveau processus stationnaire.

Preuve : voir Gourieroux et Monfort (1983), p. 148. Voir aussi Fuller (1976), chapitre 2.

Ces auteurs basent leur résultat sur le fait que si $\{X_t : t \in \mathbb{Z}\}$ est un processus stationnaire et si $\{a_i : i \in \mathbb{Z}\}$ est une suite de nombres absolument convergente, $Y_t = \sum_{i=-\infty}^{\infty} a_i X_{t-i}$ devient une expression valable au sens de la convergence en moyenne quadratique.

---

[1]. Voir Dhrymes (1974), lemme 7, p. 97 ou Gourieroux et Monfort (1983), section V.1.

Exemple

Comme exemple de processus stationnaire, prenons les bruits blancs (BB) ε_t où $\forall t \in \mathbb{Z}$ $E(\varepsilon_t) = 0$, $V(\varepsilon_t) = \sigma^2 = \lambda_0$ et $\lambda_\tau = 0$ $\forall \tau \neq 0$. Par ces propriétés, la suite $\{\varepsilon_t : t \in \mathbb{Z}\}$ est un processus stochastique stationnaire. Symboliquement, on écrit $\varepsilon_t \sim BB(0, \sigma^2)$. Soit $\{a_i : i \in \mathbb{Z}\}$ une suite absolument convergent ($\sum_{i=-\infty}^{\infty} |a_i| < \infty$). A partir des bruits blancs, on peut construire d'autres processus stationnaires. Le processus stationnaire $X_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i}$ est appelé moyenne mobile infinie.

On a finalement besoin d'une proposition concernant les polynômes formés à partir de l'opérateur retard B :

$$B^h X_t = X_{t-h} \quad , \tag{4}$$

où $B^{-1} = F$ et $FX_t = X_{t+1}$, F étant appelé l'opérateur avance.

Proposition 2

Soit le polynôme :

$$a(B) = 1 + a_1 B + a_2 B^2 \ldots + a_p B^p \quad . \tag{5}$$

Le polynôme a(B) possède un inverse unique b(B) tel que $a(B) \cdot b(B) = 1$, dès que les racines de a(B) sont de module[1] différent de 1. Si certaines racines sont de module < 1 et d'autres de module supérieur à 1, on a :

---

1. Pour $a(z) = 1 + a_1 z + a_2 z^2 \ldots + a_p z^p$, les racines de l'équation $a(z) = 0$ se notent $[z_h, h = 1, \ldots, p]$. Lorsque z_h est réelle, cette condition signifie $|z_h| \neq 1$. Pour z_h complexe ($z_h = x_h + i\, y_h$), le module de la dite racine ($|z_h| = \sqrt{x_h^2 + y_h^2}$) doit être $\neq 1$.

$b(B) = \sum_{i=-\infty}^{\infty} b_i B^i$ tel que $\sum_{i=-\infty}^{\infty} |b_i| < \infty$, avec $b_0 = 1$. Par contre, lorsque toutes les racines sont de module > 1, $b(B) = \sum_{i=0}^{\infty} b_i B^i$ tel que $\sum_{i=0}^{\infty} |b_i| < \infty$, avec $b_0 = 1$.

Preuve : voir Gourieroux et Monfort (1983), section V.2

Ceci nous permet de définir :

a) <u>processus autorégressif d'ordre p</u> AR(p)

On appelle processus autorégressif d'ordre p un processus stationnaire $\{X_t : t \in \mathbb{Z}\}$ vérifiant une relation du type :

$$(6) \qquad \sum_{j=0}^{p} a_j X_{t-j} = \varepsilon_t \quad , \qquad t \in \mathbb{Z} \quad ,$$

où ε_t est un bruit blanc de moyenne nulle et de variance σ^2.

Afin d'identifier les paramètres de la relation (6), on fixe habituellement, par convention, $a_0 = 1$. Avec l'aide du polynôme $a(B) = 1 + a_1 B + \ldots a_p B^p$, l'équation (6) peut alors s'écrire :

$$(7) \qquad a(B) X_t = \varepsilon_t \quad .$$

Avec l'aide de la proposition 2, si a(B) n'a pas de racine de module égale à 1, ce qui sera considéré comme respecté dans les pages qui suivent, le processus autorégressif d'ordre p défini par (7) s'écrit aussi de façon équivalente et unique sous forme d'une moyenne mobile infinie :

$$(8) \qquad X_t = \sum_{i=-\infty}^{\infty} b_i \varepsilon_{t-i} \quad ,$$

avec $\sum_{i=-\infty}^{\infty} |b_i| < \infty$ et $b_0 = 1$.

De plus, si les racines sont toutes de module supérieur à 1, le processus AR(p) possède une représentation moyenne mobile infinie unilatérale par l'arrière :

$$X_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \quad , \tag{8.1}$$

avec $\sum_{i=0}^{\infty} |b_i| < \infty$ et $b_0 = 1$, c'est-à-dire que X_t s'exprime comme une somme infinie convergente des ε_t passés.

Exemple

Soit le AR(1) suivant :

$$X_t = \rho X_{t-1} + \varepsilon_t \quad , \qquad t \in \mathbb{Z} \quad , \tag{9}$$

avec $|\rho| < 1$ et $\varepsilon_t \sim BB(0, \sigma^2)$. On peut vérifier que le processus $\{X_t : t \in \mathbb{Z}\}$ formé par (9) est stationnaire. En terme d'opérateur B, (9) s'écrit :

$$(1 - \rho B) X_t = \varepsilon_t \quad . \tag{10}$$

Comme $(1 - \rho z) = 0$ a pour solution $z = \frac{1}{\rho}$, et comme $|\rho| \neq 1$, d'après la proposition 2, le polynôme $(1 - \rho B)$ s'inverse. Comme $|\rho| < 1$, on a $|z| > 1$ et donc, d'après cette proposition, on aura $X_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$. Soit la somme infinie convergente[1] $\sum_{i=0}^{\infty} \rho^i B^i$, comme $\sum_{i=0}^{\infty} \rho^i B^i (1 - \rho B) = 1$, nous avons pour le AR(1), $b(B) = \sum_{i=0}^{\infty} \rho^i B^i$. Ainsi, nous avons établi que le processus AR(1) (9) avec $|\rho| < 1$ possède une représentation moyenne mobile infinie unilatérale par l'arrière.

---

[1]. On considère dans la mathématique des opérateurs que $|B|$ et $|F|$ sont < 1.

$$X_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \quad . \tag{11}$$

On n'ajoute rien si on souligne que grâce à la proposition 1, comme $\{\varepsilon_t : t \in \mathbb{Z}\}$ est stationnaire et comme $\sum_{i=0}^{\infty} |\rho^i| < \infty$ la moyenne mobile infinie (10) est nécessairement stationnaire.

Les moments de distribution du processus AR(1) peuvuent être étudiés soit par la représentation (9) soit par la représentation (11).

La démarche précédente nous a permis de garantir que si aucune racine de a(B) sont de module égal à 1, tout processus autorégressif stationnaire AR(p) peut s'exprimer comme un processus moyenne mobile infinie stationnaire. Ce fait incite certains auteurs à voir la stationnarité d'un processus comme suit : lorsque X_t peut-être exprimé comme une somme convergente sur la série stationnaire $\{\varepsilon_t : t \in \mathbb{Z}\}$, le processus $\{X_t\}$ est lui-même stationnaire. C'est selon cet angle que nous approcherons le problème de la stationnarité spatiale.

En manipulant l'expression (6), on peut considérer l'exemple suivant[1] :

$$X_t = a(X_{t-1} + X_{t+1}) + \varepsilon_t \quad , \quad t \in \mathbb{Z} \quad , \tag{12}$$

$$\varepsilon_t \sim BB(0, \sigma^2).$$

---

[1]. Whittle (1954) appelle ce modèle : "Line Transect". Il est utile lors de l'étude de processus spatiaux en 1 dimension. Il est aussi étudié par Haining (1978b).

D'après Haining (1978b), lorsque $|a| < 1/2$, ce modèle est stationnaire et X_t a alors une représentation MA(∞) bilatérale [double infinie] stationnaire. Pour se convaincre, en exprimant en terme d'opérateur l'expression (12), on obtient :

$$[1 - a(B + F)]\, X_t = \varepsilon_t \quad , \tag{13}$$

où $B^{-1} = F$ et $F \cdot X_t = X_{t+1}$.

Comme unique solution MA(∞) de (13), il obtient :

$$X_t = \frac{\varepsilon_t}{1 - a(B + F)} = \sum_{r=0}^{\infty} \sum_{k=0}^{r} \binom{r}{k} a^r \varepsilon_{t+2k-r} \quad . \tag{14}$$

En terme de la proposition 1, cette équation est valable au sens de la convergence en moyenne quadratique lorsque $|a| < 1/2$. Pour obtenir l'équation à l'extrême droite, il suffit d'appliquer la formule d'expansion binomiale suivante :

$$(a + b)^r = \sum_{k=0}^{r} \binom{r}{k} a^{r-k} b^k \quad . \tag{15}$$

Ainsi, par $\frac{1}{(1 - d)} = \sum_{j=0}^{\infty} d^j$, pour $d < 1$,

$$X_t = \sum_{r=0}^{\infty} a^r (B + F)^r \varepsilon_t \quad ,$$

$$= \sum_{r=0}^{\infty} a^r \sum_{k=0}^{r} \binom{r}{k} B^{r-k} F^k \varepsilon_t \quad ,$$

$$X_t = \sum_{r=0}^{\infty} a^r \sum_{k=0}^{r} \binom{r}{k} \varepsilon_{t+2k-r} \quad . \tag{16}$$

Pour que l'équation

$$X_t = \sum_{r=0}^{\infty} \sum_{k=0}^{r} \binom{r}{k} a^r \varepsilon_{t+2k-r} \qquad (17)$$

soit valable au sens de la convergence en moyenne quadratique, il suffit de vérifier sous quelles conditions $\sum_{r=0}^{\infty} \sum_{k=0}^{r} \binom{r}{k} |a|^r < \infty$. Le fait que $\forall a$, $\sum_{k=0}^{r} \binom{r}{k} |a|^r = (2|a|)^r$ et que pour $|a| < 1/2$, $\sum_{r=0}^{\infty} (2a)^r = \frac{1}{1-2a} < \infty$, vérifie la condition de stationnarité $|a| < 1/2$ du modèle (12).

Comme $\sum_{r=0}^{\infty} \sum_{k=0}^{r} \binom{r}{k} |a^r| = \sum_{r=0}^{\infty} \sum_{k=0}^{r} |\binom{r}{k}^{1/r} a|^r$ implique[1] que $\sum_{r=0}^{\infty} \sum_{k=0}^{r} [\binom{r}{k}^{2/r} a^2]^r = \sum_{r=0}^{\infty} \sum_{k=0}^{r} \binom{r}{k}^2 a^{2r} < \infty$, grâce à $E(X_t) = 0$, on obtient directement que pour $|a| < 1/2$, $V(X_t) = E(X_t^2) = \sigma^2 \sum_{r=0}^{\infty} \sum_{k=0}^{r} \binom{r}{k}^2 a^{2r} < \infty$.

b) <u>Processus moyenne mobile d'ordre q</u> MA(q)

On appelle processus moyenne mobile d'ordre q un processus $\{X_t : t \in \mathbb{Z}\}$ défini par :

$$X_t = \sum_{j=0}^{q} b_j \varepsilon_{t-j} \quad , \qquad t \in \mathbb{Z} \quad , \qquad (18)$$

où $\varepsilon_t \sim BB(0, \sigma^2)$ et où, pour l'identification des paramètres, on pose $b_0 = 1$. En terme du polynôme $b(B) = (1 + b_1 B + b_2 B^2 + \ldots + b_q B^q)$, on a :

$$X_t = b(B)\, \varepsilon_t \quad .$$

---

[1]. On utilise ici la propriété $\sum_{i=-\infty}^{\infty} |a_i| < \infty \rightarrow \sum_{i=-\infty}^{\infty} a_i^2 < \infty$. Voir Fuller (1976), p. 27.

Evidemment, comme en (18) la somme est sur un nombre fini (q) d'éléments, un processus MA(q) est toujours stationnaire. La dualité du MA(q) avec le AR(p) s'exprime comme suit.

Grâce à la proposition 2, si b(B) n'a pas de racine de module égal à 1, le processus (18) s'écrit de façon équivalente comme un AR(∞) stationnaire :

$$(19) \qquad \sum_{i=-\infty}^{\infty} a_i X_{t-i} = \varepsilon_t$$

avec $\sum_{i=\infty}^{\infty} |a_i| < \infty$ et $a_0 = 1$.

De plus, si les racines sont toutes de module supérieur à 1, la représentation AR(∞) est unilatérale par l'arrière,

$$(19.1) \qquad \sum_{i=0}^{\infty} a_i X_{t-i} = \varepsilon_t$$

avec $\sum_{i=0}^{\infty} |a_i| < \infty$ et $a_0 = 1$.

Un processus MA(q) exprimable sous forme AR(∞) stationnaire est dit inversible. Pour d'autres propriétés, consulter Gourieroux et Monfort (1983), section V.B.

1.3) Concepts d'analyse spectrale

Comme approche alternative au domaine temporel, il existe l'analyse spectrale[1]. Comme Granger et Newbold (1976) le soulignent, page 44, lorsqu'un résultat peut être obtenu dans un de ces deux domaines, il peut l'être dans l'autre. Il est souvent plus facile de l'obtenir en analyse spectrale. On verra plus loin que ceci est aussi vrai en analyse des champs aléatoires. Mais avant :

soit $\{X_t\}$ un processus stationnaire de moyenne nulle avec autocovariance $\lambda_\tau = \text{cov}\,(X_t, X_{t-\tau})$ et autocorrélation $\rho_\tau = \frac{\lambda_\tau}{\lambda_o}$. La séquence des autocovariances est résumée par la fonction génératrice d'autocovariances[2] :

$$\lambda(z) = \sum_{\tau=-\infty}^{\infty} \lambda_\tau z^\tau \quad , \tag{20}$$

où $\lambda_\tau = \lambda_{-\tau}$.

Relié à (20) où l'on remplace z par $e^{-i\omega}$ où $i = \sqrt{-1}$, il existe la fonction $s(\omega)$ appellée fonction de densité spectrale, spectre de puissance ou représentation de Fourier de λ_τ. On écrit :

$$s(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \lambda_\tau e^{-i\tau\omega} = \frac{\lambda_o}{2\pi} + \frac{1}{\pi} \sum_{\tau=1}^{\infty} \lambda_\tau \cos(\tau\omega) \quad , \tag{21}$$

$s(\omega)$ est une fonction réelle non négative de période 2π, c'est-à-dire que $s(\omega)$ se reproduit en dehors du domaine $-\pi \leq \omega \leq \pi$. On a ainsi :

$s(\omega + 2\pi k) = s(\omega)$ $\forall k$ entier.

Comme $\int_{-\pi}^{\pi} e^{i(k-\tau)\omega}\, d\omega = \begin{cases} 0 & \text{si } k \neq \tau \\ 2\pi & \text{si } k = \tau \end{cases}$

---

[1]. Voir Fuller (1976). chapitres 3 et 4, Gourieroux et Monfort (1983), chapitre VII, ou Malinvaud (1978), chapitre II.

[2]. Voir Granger et Newbold, chapitre 1.

par évaluation directe, on a :

$$(22) \qquad \lambda_\tau = \int_{-\pi}^{\pi} e^{i\tau\omega} \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \lambda_k e^{-ik\omega} d\omega = \int_{-\pi}^{\pi} e^{i\tau\omega} s(\omega) d\omega \quad .$$

A cause de la relation inverse entre (22) et (21), (22) est dit : représentation de Fourier inverse de $s(\omega)$ ou représentation spectrale des autocovariances. A partir de (22), la représentation spectrale des autocorrélations s'écrit :

$$(23) \qquad \rho_\tau = \frac{\lambda_\tau}{\lambda_o} = \int_{-\pi}^{\pi} e^{i\tau\omega} \frac{s(\omega)}{\lambda_o} d\omega \quad .$$

Comme $s(\omega) \geq 0$ et comme d'après (23) $1 = \rho_o = \int_{-\pi}^{\pi} \frac{s(\omega)}{\lambda_o} d\omega$, $\frac{s(\omega)}{\lambda_o}$ est une fonction qui, sur le domaine $[-\pi, \pi]$ a toutes les propriétés d'une fonction de densité. Soit $S(\omega)$ la fonction de distribution spectrale (mesure spectrale) où $S(\omega) = \int_{-\pi}^{\omega} \frac{s(\omega')}{\lambda_o} d\omega'$ telle que $S(-\pi) = 0$, $S(\pi) = 1$ et $S(0) = 1/2$; (23) s'écrit alors de façon équivalente :

$$(23.1) \qquad \rho_\tau = \frac{\lambda_\tau}{\lambda_o} = \int_{-\pi}^{\pi} e^{i\tau\omega} dS(\omega) \quad ,$$

car en effet, $dS(\omega) = d \int_{-\pi}^{\pi} \frac{s(\omega')}{\lambda_o} d\omega' = \frac{s(\omega)}{\lambda_o} d\omega$.

Comme autre relation fondamentale, il existe la représentation spectrale d'une série stationnaire non déterministe (représentation de Cramer),

$$(24) \qquad X_t = \int_{-\pi}^{\pi} e^{i\tau\omega} dz_x(\omega) \ ^1 \ ,$$

où

$$(24.1) \qquad E(dz_x(\omega) \cdot \overline{dz_x(\gamma)}) \quad \begin{cases} = 0 & \omega \neq \gamma \\ = s(\omega)\, d\omega & \omega = \gamma \end{cases} \ ,$$

---

[1]. Pour des détails sur la variable aléatoire dz_x, consulter particulièrement Gourieroux et Monfort (1983), chapitre VII.

où $\overline{z(\gamma)}$ désigne la fonction orthogonale complexe conjuguée de $z(\gamma)$. Comme Granger et Newbold le soulignent, toute série stationnaire non déterministe peut être exprimée selon la forme (24). On peut vérifier que la condition (24.1) est nécessaire pour que X_t soit stationnaire[1]. On va maintenant présenter la densité spectrale (21) pour les processus stochastiques les plus communs.

a) Spectre d'un bruit blanc

Par $E(\varepsilon_t) = 0 \quad \forall t \in \mathbb{Z}$

et $$E(\varepsilon_t \varepsilon_s) = \begin{cases} \sigma^2 & t = s \\ 0 & t \neq s \end{cases}$$

alors

$$s_\varepsilon(\omega) = \frac{\sigma^2}{2\pi} . \tag{25}$$

b) Une propriété du spectre

Soit X_t une série stationnaire avec spectre $s_x(\omega)$. Le spectre de :

$$Y_t = \sum_{j=0}^{m} c_j X_{t-j} = c(B) X_t$$

s'écrit :

$$s_y(\omega) = c(e^{-i\omega}) \cdot c(e^{i\omega}) \cdot s_x(\omega) . \tag{26}$$

---

1. Soit $\overline{X_{t-\tau}} = \int_{-\pi}^{\pi} e^{-i(t-\tau)} \overline{dz_x(\gamma)}$ et $E[dz_x(\omega)] = E[dz_x(\gamma)] = 0$, ce qui implique par (24) que $E(X_t) = 0$. Comme,

$\lambda_\tau = E[X_t \cdot \overline{X_{t-\tau}}] = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{it\omega} e^{-i(t-\tau)\gamma} E[dz_x(\omega) \cdot \overline{dz_x(\gamma)}]$. Cette expression est égale à (22) que lorsque (24.1) est respecté.

La relation (26) est très importante lorsqu'on veut obtenir le spectre d'un AR(p) ou d'un MA(q).

c) Spectre d'un MA(1)

$$X_t = \varepsilon_t + b\,\varepsilon_{t-1} = (1 + bB)\,\varepsilon_t \quad .$$

On a donc :

$$(27) \qquad s_x(\omega) = \frac{\sigma^2}{2\pi}\,(1 + b\,e^{i\omega})\,(1 + b\,e^{-i\omega}) \ ,$$

$$s_x(\omega) = \frac{\sigma^2}{2\pi}\,(1 + 2\,b\cos(\omega) + b^2) \quad .$$

Par (22), on peut évidemment retrouver, grâce à (27), la série des autocovariances.

Exemple

$$(28) \qquad \lambda_o = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}(1 + 2\,b\cos(\omega) + b^2)\,d\omega \quad .$$

On a évidemment, comme attendu[1] :

$$(29) \qquad \lambda_o = \sigma^2\,(1 + b^2) \quad .$$

Présentons maintenant les modifications qu'il faut apporter à ces concepts lorsqu'on ajoute une dimension.

---

[1]. $\lambda_o = \frac{\sigma^2}{2\pi}\,(1 + b^2)\int_{-\pi}^{\pi} 1\,d\omega + \frac{\sigma^2 b}{\pi}\int_{-\pi}^{\pi}\cos(\omega)$

$= \frac{\sigma^2}{2\pi}\,(1 + b^2)\cdot 2\pi + 0 \qquad$ car $\int_{-\pi}^{\pi}\cos(\omega) = \sin(\omega)\Big|_{-\pi}^{\pi} = 0.$

2) Modèles autorégressifs spatiaux

2.1) Deux modèles autorégressifs spatiaux stationnaires

Comme remarqué par Granger (1969), les outils de la série temporelle peuvent être étendus à la modélisation des séries spatiales. On reviendra plus loin à la critique que fait cet auteur sur cette généralisation qui, du point de vue mathématique est correcte mais qui, du côté interprétation économique, est discutable et non réaliste. Mais avant, généralisons les concepts des sections précédentes.

Le modèle AR(p) temporel devient pour le cas spatial, le modèle AR(M,N)

$$(30) \qquad \sum_{s=0}^{M} \sum_{t=0}^{N} a_{s,t} X_{j-s,k-t} = u_{j,k} \qquad j,k \in \mathbb{Z}^2$$

où $a_{0,0} = 1$, $a_{M,N} \neq 0$, $a_{M,0} \neq 0$, $a_{0,N} \neq 0$, et $\{u_{j,k}\}$ est une séquence de variables aléatoires (v.a.) non correlées de moyenne nulle et de variance finie, que l'on note : $u_{j,k} \sim BB(0, \sigma^2)$.

Soit $B_j X_{j,k} = X_{j-1,k}$ et $B_k X_{j,k} = X_{j,k-1}$, l'expression (30) devient :

$$(31) \qquad \sum_{s=0}^{M} \sum_{t=0}^{N} a_{s,t} B_j^s B_k^t X_{j,k} = u_{j,k} \ ,$$

ou encore, en terme du polynôme $a(B_j,B_k) = \sum_{s=0}^{M} \sum_{t=0}^{N} a_{s,t} B_j^s B_k^t$,

$$(32) \qquad a(B_j, B_k) X_{j,k} = u_{j,k} \ .$$

Haining (1978b) montre que, pourvu qu'aucune racine de $a(B_j, B_k)$ ne soit de module égal à 1,

$$(33) \qquad X_{j,k} = \frac{u_{j,k}}{a(B_j, B_k)} = \sum_p \sum_q b_{p,q}\, u_{j-p,k-q}$$

est l'unique représentation MA(∞,∞) de la série $\{X_{j,k}\}$ laquelle est valide au sens de la convergence en moyenne quadratique. En effet, par les représentations spectrales de $u_{j,k}$ et $X_{j,k}$ qui s'écrivent :

$$(34) \qquad u_{j,k} = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{ij\omega_1} e^{ik\omega_2}\, dz_u\,(\omega_1, \omega_2)$$

et

$$(35) \qquad X_{j,k} = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{ij\omega_1} e^{ik\omega_2}\, dz_x\,(\omega_1, \omega_2)$$

où $z_u\,(\omega_1, \omega_2)$ et $z_x\,(\omega_1, \omega_2)$ sont des fonctions orthogonales telles que[1] :

$$(36) \qquad \frac{\partial\, z_x\,(\omega_1, \omega_2)}{\partial\, z_u\,(\omega_1, \omega_2)} = \left[a(e^{-i\omega_1}, e^{-i\omega_2})\right]^{-1} .$$

Haining (1978b) obtient que l'unique solution stationnaire de (30) s'écrit grâce à (36) et (35) :

$$(37) \qquad X_{j,k} = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{ij\omega_1} e^{ik\omega_2} \left[a(e^{-i\omega_1}, e^{-i\omega_2})\right]^{-1} dz_u\,(\omega_1, \omega_2) ,$$

---

[1]. Ici, on met $B_j = e^{-i\omega_1}$ et $B_k = e^{-i\omega_2}$. Remarquons de plus que le spectre de : $X_{j,k} = a(B_j, B_k)\, u_{j,k}$, où $s_u\,(\omega_1, \omega_2) = \frac{\sigma^2}{4\pi^2}$, sera $s_x\,(\omega_1, \omega_2) = \frac{\sigma^2}{4\pi^2}\, a\,(e^{-i\omega_1}, e^{-i\omega_2})\, a\,(e^{i\omega_1}, e^{i\omega_2})$.

ce qui devient après le développement de Fourier de $\left[a(e^{-i\omega_1}, e^{-i\omega_2})\right]^{-1}$:

$$X_{jk} = \sum_p \sum_q b_{p,q}\, u_{j-p,k-q} \quad .$$

Exemple

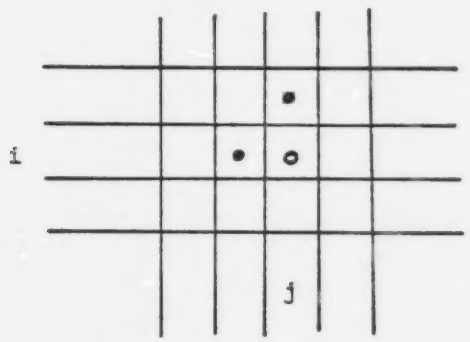
Le modèle AR(1, 1) peut s'écrire :

$$X_{i,j} = \alpha(X_{i-1,j} + X_{i,j-1}) + u_{i,j} \qquad i,j \in \mathbb{Z}^2 \tag{38}$$

$$\text{avec } u_{i,j} \sim BB(0, \sigma^2)$$

où on pose ici $a_{1,0} = a_{0,1} = \alpha$. Haining appelle ce modèle, le "wave model", car en fait, comme une vague, l'observation $X_{i,j}$ ne dépend que de l'arrière. Le schéma de dépendance se retrouve à la figure 1. C'est en fait le modèle temporel (dépendance unilatérale) appliqué à une version spatiale[1].

FIGURE 1



[1]. Comme Haining (1978b) le souligne, dans un contexte agricole, la production d'un arbre fruitier dans un champ à forte pente ou à fort vent à direction prédominante peut être caractérisé par ce genre de modèle.

Ce modèle autorégressif si $|\alpha| < 1/2$ sera stationnaire au sens de la convergence en moyenne quadratique. En terme du polynôme $a(B_i, B_j)$, on a :

$$[1 - \alpha(B_i + B_j)] X_{i,j} = u_{i,j} \quad . \tag{39}$$

D'après (33), l'unique représentation stationnaire MA(∞,∞) est :

$$X_{i,j} = \frac{u_{i,j}}{[1 - \alpha(B_i + B_j)]} \quad , \tag{40}$$

ce qui se réécrit, on le sait :

$$X_{i,j} = \sum_{r=0}^{\infty} \alpha^r (B_i + B_j)^r u_{i,j} \quad , \tag{41}$$

$$X_{i,j} = \sum_{r=0}^{\infty} \alpha^r \sum_{t=0}^{r} \binom{r}{t} B_i^{r-t} B_j^t u_{i,j} \quad . \tag{42}$$

Posons $s = r-t \rightarrow r = s+t$. On obtient alors :

$$X_{i,j} = \sum_{s+t=0}^{\infty} \sum_{t=0}^{s+t} \alpha^{s+t} \binom{s+t}{t} B_i^s B_j^t u_{i,j} = \sum_{s+t=0}^{\infty} \sum_{t=0}^{s+t} \alpha^{s+t} \binom{s+t}{t} u_{i-s,j-t} \quad , \tag{43}$$

et on a donc une représentation MA(∞, ∞) unilatérale vers l'arrière.

Pour que cette expression soit valable au sens de la convergence en moyenne quadratique, comme $u_{i,j} \sim BB(0, \sigma^2)$, il faut que :

$$\sum_{s+t=0}^{\infty} \sum_{t=0}^{s+t} \binom{s+t}{t} |\alpha|^{s+t} < \infty \quad ,$$

ce qui s'écrit :

$$(44) \qquad \sum_{s+t=0}^{\infty} (2|\alpha|)^{s+t} = \frac{1}{1-2|\alpha|} < \infty \quad ,$$

lorsque $|\alpha| < 1/2$. Comme d'habitude (44) implique que :

$$\sum_{s+t=0}^{\infty} \sum_{t=0}^{s+t} \binom{s+t}{t}^2 \alpha^{2(s+t)} < \infty$$

et donc $E(X_{i,j}^2) = \sigma^2 \sum_{s+t}^{\infty} \sum_{t=0}^{s+t} \binom{s+t}{t}^2 \alpha^{2(s+t)} < \infty$. Voir Haining (1978b), p. 32-3.

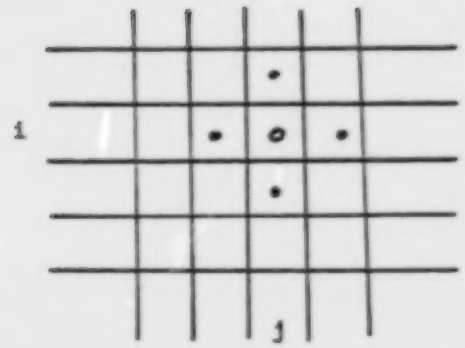
On vient donc d'établir que lorsque $|\alpha| < 1/2$, le modèle (38) possède une représentation MA(∞, ∞) unilatérale stationnaire.

En plus de s'intéresser au modèle en vague de l'équation (38), Haining étudie les conditions de stationnarité du modèle quadrilatéral autorégressif (Q.A.) suivant, dont le schéma de dépendance se retrouve à la figure 2.

$$(45) \qquad X_{i,j} = \rho(X_{i-1,j} + X_{i+1,j} + X_{i,j-1} + X_{i,j+1}) + u_{i,j} \quad , \quad i,j \in \mathbb{Z}^2 \quad ,$$

$$u_{i,j} \sim BB(0, \sigma^2) \quad .$$

FIGURE 2



Lorsque $|\rho| < 1/4$, ce modèle autorégressif est stationnaire.

Une expression algébrique de la représentation MA(∞, ∞) bilatérale stationnaire n'ayant jamais été présentée dans la littérature, on juge bon de le faire ici. On pourra de plus vérifier que lorsque $|\rho| < 1/4$, le modèle est stationnaire au sens de la convergence en moyenne quadratique.

Soit $F_i = B_i^{-1}$ et $F_j = B_j^{-1}$, en terme du polynôme, la solution MA(∞, ∞) de (45) devient :

$$X_{i,j} = \frac{u_{i,j}}{[1 - \rho(B_i + F_i + B_j + F_j)]} \cdot$$

$[1 - \rho([B_i + B_j] + [F_i + F_j])]^{-1}$ s'écrit[1] :

$$= \sum_{r=0}^{\infty} \rho^r [(B_i + B_j) + (F_i + F_j)]^r = \sum_{r=0}^{\infty} \rho^r \sum_{t=0}^{r} \binom{r}{t} (B_i + B_j)^{r-t} (F_i + F_j)^t \quad ,$$

---

[1]. Au lieu de faire le développement qui suit, nous aurions tout aussi bien pu utiliser la formule du quadrinôme.

Et en posant $s = r - t \rightarrow s + t = r$

$$= \sum_{s+t=0}^{\infty} \rho^{s+t} \sum_{t=0}^{s+t} \binom{s+t}{t} (B_i + B_j)^s (F_i + F_j)^t$$

$$= \sum_{s+t=0}^{\infty} \rho^{s+t} \sum_{t=0}^{s+t} \binom{s+t}{t} \sum_{v=0}^{s} \binom{s}{v} B_i^{s-v} B_j^v \sum_{w=0}^{t} \binom{t}{w} F_i^{t-w} F_j^w \quad .$$

Posons $y = t - w \rightarrow t = y + w$

$z = s - v \rightarrow s = z + v$

et pour simplifier la notation, écrivons : $\tau = v + w + y + z$

$$= \sum_{\tau=0}^{\infty} \rho^{\tau} \sum_{y+w=0}^{\tau} \binom{\tau}{y+w} \sum_{v=0}^{z+v} \binom{z+v}{v} B_i^z B_j^v \sum_{w=0}^{y+w} \binom{y+w}{w} F_i^y F_j^w \quad .$$

On a donc :

$$(46) \qquad X_{i,j} = \sum_{\tau=0}^{\infty} \sum_{y+w=0}^{\tau} \sum_{v=0}^{z+v} \sum_{w=0}^{y+w} \rho^{\tau} \binom{\tau}{y+w} \binom{z+v}{v} \binom{y+w}{w} u_{i+y-z,\, j+w-v} \quad ,$$

et on vérifie bien la dépendance bilatérale de cette représentation $MA(\infty, \infty)$.

Pour s'assurer que cette quadruple somme infinie est sommable de façon absolue lorsque $|\rho| < 1/4$, écrivons :

$$= \sum_{\tau=0}^{\infty} \sum_{y+w=0}^{\tau} \binom{\tau}{y+w} \underbrace{\sum_{v=0}^{z+v} |\rho|^{z+v} \binom{z+v}{v}}_{(2|\rho|)^{z+v}} \cdot \underbrace{\sum_{w=0}^{y+w} \binom{y+w}{w} |\rho|^{y+w}}_{(2|\rho|)^{y+w}}$$

$$= \sum_{\tau=0}^{\infty} \sum_{y+w=0}^{\tau} \binom{\tau}{y+w} \cdot (2|\rho|)^{\tau}$$

$$= \sum_{\tau=0}^{\infty} (4|\rho|)^{\tau} = \frac{1}{1 - 4|\rho|} < \infty \qquad \text{si } |\rho| < 1/4 \; .$$

On a donc vérifié que lorsque $|\rho| < 1/4$, le processus spatial autorégressif (45) est stationnaire et possède une représentation MA(∞, ∞) bilatérale stationnaire.

En se limitant à ces deux modèles autorégressifs qui permettent de générer un champ aléatoire $\{X_{i,j} : i,j \in \mathbb{Z}^2\}$ stationnaire, Haining veut savoir si, tout comme en série chronologique, on peut à partir des autocorrélations empiriques, discriminer entre ces deux modèles, c'est-à-dire, est-ce que le modèle (45) donnera un tableau de corrélation assez différent de celui de (38), pour qu'on puisse identifier lequel des deux modèles a généré notre série spatiale observée? Mais avant de passer aux détails, on aimerait définir le concept d'isotropie. Pour ce faire, on étudiera ici le modèle (Q.M.) quadrilatéral moyenne mobile développé par Haining (1978c).

2.2) Modèle spatial moyenne mobile et concept d'isotropie

On a décidé de définir le concept d'isotropie à partir du modèle Q.M. car les corrélations sont beaucoup plus faciles à obtenir qu'avec le modèle autorégressif Q.A.

Haining (1978c) développe les conditions pour qu'un MA(M, N)[1] possède une représentation AR(∞, ∞) stationnaire. Il faut que le processus MA(M,N) soit inversible, c'est-à-dire que la dualité entre le AR(p) et MA(q) de la série chronologique se retrouve aussi dans le domaine spatial. De fait, les conditions d'inversibilité du MA(M, N) sont les mêmes que pour la stationnarité du AR(M, N). Pour plus de détail, consultez cette référence. Le modèle Q.M. s'écrit :

---

[1]. Comme M et N sont finis, le processus MA(M, N) est évidemment toujours stationnaire.

$$(47) \qquad X_{i,j} = \theta (u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}) + u_{i,j} , \quad i,j \in \mathbb{Z}^2 ,$$

$$u_{i,j} \sim BB(0, \sigma^2) \quad .$$

Si $|\theta| < 1/4$, le modèle Q.M. est inversible en modèle AR(∞, ∞) bilatéral stationnaire infini. La version non contrainte entre les paramètres s'écrit :

$$(48) \qquad X_{i,j} = \alpha u_{i-1,j} + \beta u_{i+1,j} + \gamma u_{i,j-1} + \delta u_{i,j+1} + u_{i,j} .$$

Le tableau des autocorrélations $\rho(s,t)$ ($s,t = 0, \pm 1, \pm 2, \pm 3, \ldots$) peut être construit par la méthode directe habituellement employée en série chronologique, ou bien grâce à la fonction de densité spectrale du modèle (48), qui s'écrit :

$$(49) \qquad s_x(\omega_1, \omega_2) = \frac{\sigma^2}{4\pi^2} (1 + \alpha e^{-i\omega_1} + \beta e^{i\omega_1} + \gamma e^{-i\omega_2} + \delta e^{i\omega_2}).$$
$$(1 + \alpha e^{i\omega_1} + \beta e^{-i\omega_1} + \gamma e^{i\omega_2} + \delta e^{-i\omega_2}) \quad .$$

$\sigma(s,t)$ devient alors le coefficient associé à $e^{is\omega_1} \cdot e^{it\omega_2}$ dans le développement de (49). La séquence des autocorrélations $\rho(s,t) = \frac{\sigma(s,t)}{\sigma(0,0)}$ se retrouve au tableau 1 suivant :

TABLEAU 1

Autocorrélations du modèle (48)

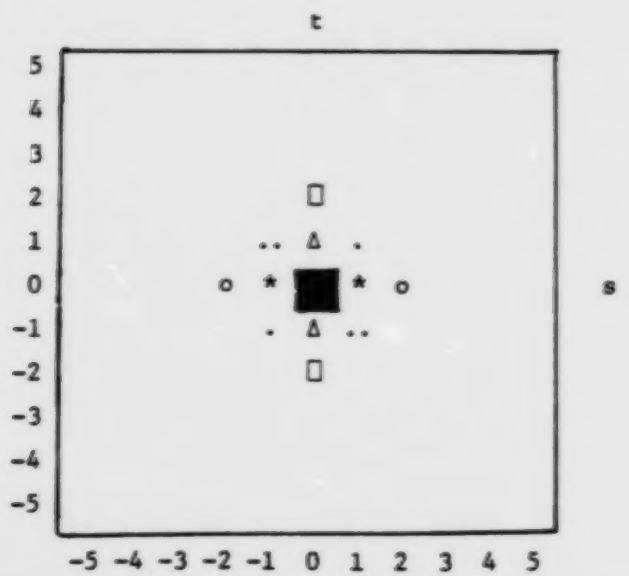
$\rho(0,1) = \rho(0,-1) = \frac{\delta+\gamma}{K}$	*	$\rho(1,0) = \rho(-1,0) = \frac{\alpha+\beta}{K}$	Δ
$\rho(0,2) = \rho(0,-2) = \frac{\gamma\cdot\delta}{K}$	o	$\rho(2,0) = \rho(-2,0) = \frac{\alpha\cdot\beta}{K}$	$\square$
$\rho(1,1) = \rho(-1,-1) = \frac{\alpha\delta+\beta\gamma}{K}$	.	$\rho(-1,1) = \rho(1,-1) = \frac{\alpha\gamma+\beta\delta}{K}$	..
$\rho(s,t) = 0 \ (\lvert s\rvert+\lvert t\rvert) > 2$		$K = (1+\alpha^2+\beta^2+\gamma^2+\delta^2)$	

source : Haining (1978c)

On remarque que, tout comme en série chronologique, le modèle MA possède des autocorrélations qui deviennent nulles après un nombre limité de délais.

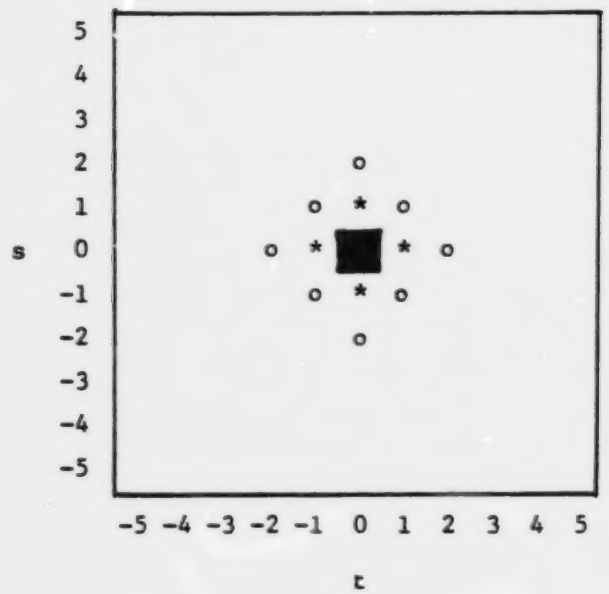
Si on place les sigles du tableau 1 sur la figure 3 suivante, on observe la présence de tendances directionnelles dans les autocorrélations.

FIGURE 3



La symétrie des autocorrélations $\rho(s,t) = \rho(-s,-t)$ explique pourquoi les auteurs ne présentent habituellement que les quadrants 1 et 4 du tableau des corrélations. Par rapport aux axes $s = 0$ et $t = 0$, on constate qu'un champ aléatoire stationnaire génère des corrélations qui dépendent des directions. Un champ aléatoire isotrope n'aura pas ce biais directionnel dans les corrélations. C'est pourquoi, comme on peut le noter en posant $\alpha = \beta = \delta = \gamma = \theta$, le modèle Q.M. appartient à la classe des modèles isotropes. En effet, la structure des autocorrélations du modèle (47) produit la figure 4 suivante :

FIGURE 4



où on a pour un modèle isotrope rectangulaire :

$$\rho(s,t) = \rho(s,-t) = \rho(-s,t) = \rho(-s,-t) \quad .$$

Donc, un champ stochastique est isotrope au sens large lorsque les corrélations sont invariantes par rapport à la direction. Voir Cliff et Ord (1975a) ou Haining (1980).

Ainsi le modèle :

$$X_{i,j} = \rho(X_{i-1,j} + X_{i+1,j} + X_{i,j-1} + X_{i,j+1}) + u_{i,j} \quad (50)$$

est isotrope, alors que le modèle :

$$X_{i,j} = a(X_{i-1,j} + X_{i+1,j}) + b(X_{i,j-1}, X_{i,j+1}) + u_{i,j} \quad (51)$$

ne l'est pas. Par exemple, si $a > b$, les corrélations sur l'axe (s) nord-sud seront plus fortes que sur l'axe est-ouest (t).

2.3) La spécification dans les modèles autorégressifs spatiaux

De même que la procédure en 3 étapes de la série chronologique, on voudrait ici, à partir du tableau des corrélations spatiales empiriques R(s,t), pouvoir identifier le modèle théorique qui a pu générer la série observée.

Le problème est toutefois plus complexe qu'en séries chronologiques. En effet, la classe de modèles théoriques spatiaux autorégressifs stationnaires est beaucoup plus large qu'en séries temporelles. En plus du problème habituel de spécification de l'ordre de dépendance du schéma AR, MA ou ARMA, s'ajoute ici celui de l'existence de directions de dépendances dans la construction du modèle, c'est-à-dire qu'on doit se demander si la structure de dépendance est bilatérale ou unilatérale et pour chacun de ces cas, est-ce que la formulation est isotrope ou non.

Par exemple, les 2 versions isotropes les plus triviales de l'équation (30) sont :

$$X_{i,j} = \alpha(X_{i-1,j} + X_{i,j-1}) + u_{i,j} \quad , \tag{52}$$

(version unilatérale)

et

$$X_{i,j} = \rho(X_{i-1,j} + X_{i+1,j} + X_{i,j-1} + X_{i,j+1}) + u_{i,j} \quad . \tag{53}$$

(version bilatérale)

Comme modèle de base non isotrope, on pourrait avoir :

$$X_{i,j} = a\, X_{i-1,j} + b\, X_{i,j-1} + u_{i,j} \tag{54}$$

et

$$X_{i,j} = c\,(X_{i-1,j} + X_{i+1,j}) + d(X_{i,j-1} + X_{i,j+1}) + u_{i,j} \quad . \tag{55}$$

La multiplicité de modèles autorégressifs qui originent de (30) montre la complexité que peut prendre la phase de spécification dans le domaine spatial. Avant d'établir les caractéristiques théoriques des corrélogrammes de deux de ces modèles, présentons, en premier lieu, les outils empiriques nécessaires.

2.3.1) *Le corrélogramme spatial empirique*

Comme

$$\rho(s,t) = \frac{\text{cov}\,(X_{i,j},\, X_{i+s,j+t})}{\text{var}\,(X_{i,j})} \quad , \tag{56}$$

en nous limitant à notre série $x_{(PQ\times 1)}$ (réalisation de $X_{(PQ\times 1)}$) observée sur le plan rectangulaire P×Q (P zones sur la direction nord-sud et Q sur la direction est-ouest), l'estimateur naturel de (56) est :

$$(57) \qquad R(s,t) = \frac{\frac{1}{(P-s)(Q-t)} \sum_{i=1}^{P-s} \sum_{j=1}^{Q-t} z_{i,j}\, z_{i+s,j+t}}{\frac{1}{PQ} \sum_{t=1}^{P} \sum_{s=1}^{Q} z_{i,j}^2}$$

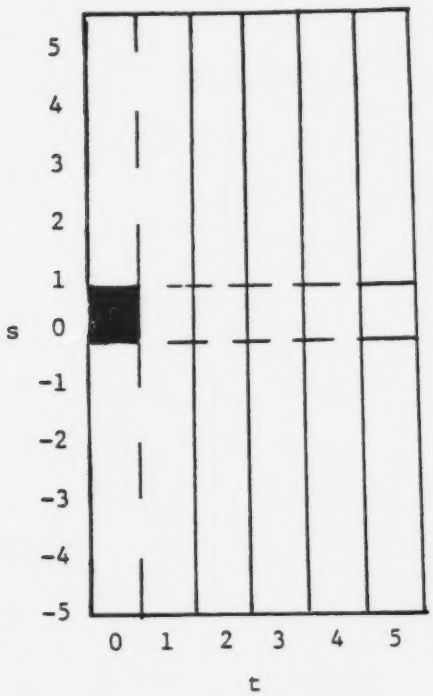
où $\bar{x} = \dfrac{\sum_{i=1}^{P} \sum_{j=1}^{Q} x_{ij}}{PQ}$ et $z_{i,j} = (x_{i,j} - \bar{x})$.

Ceci peut être approximé, dans de grands échantillons par :

$$R(s,t) = \frac{\sum_{i=1}^{P-s} \sum_{j=1}^{Q-t} z_{i,j} \cdot z_{i+s,j+t}}{\sum_{i=1}^{P} \sum_{j=1}^{Q} z_{i,j}^2} \quad .$$

A partir du tableau des corrélations empiriques, qui se limite aux quadrants 1 et 4, pour t et s ε [-5,...,+5], ce tableau a la forme suivante :

FIGURE 5



Si on prend comme référence, par exemple, les axes $s = 0$ et $t = 0$, le corrélogramme estimé peut nous fournir de l'information sur la présence de tendances directionnelles dans les corrélations (indication de formulation non isotrope). Si les corrélations sont assez fortes sur les deux axes (nord-sud pour s et est-ouest pour t), on peut croire à une formulation bilatérale. Dans le cas où, par exemple, dans les deux sens, les corrélations semblent fortes de façon inégale, les modèles (54) ou (55) peuvent être postulés. Si les corrélations sur l'axe t sont presque nulles, cela peut suggérer les modèles (54) ou (55) avec $b = d = 0$. Un tableau d'auto-corrélation empirique où les quadrants 1 et 4 se ressemblent nous porte à choisir des versions isotropes comme (52) et (53) [voir Whittle (1954)].

Toutefois, comment faire pour discerner entre les modèles (52) et (53) ou (54) et (55). Haining (1978b), chapitre 3, s'intéresse à savoir si, à partir des autocorrélations empiriques, on peut faire un choix entre le modèle non symétrique (52) et la version symétrique (53). C'est ce qui va nous intéresser à la section 2.3.2). Mais avant, on aimerait faire quelques remarques sur la phase de spécification en analyse spatiale.

Remarques

1) La loi de distribution jointe des autocorrélations empiriques n'est pas encore développée. Il devient alors difficile de faire des tests de signification sur les autocorrélations R(s,t). Il faut donc pour l'instant se contenter d'une méthode visuelle[1].

2) Dans la même ligne d'idée, aucun test formel théorique n'a été présenté pour tester la présence de biais directionnel dans les autocorrélations. Pour une tentative intéressante, voir Gatrell (1979b).

3) Pour utiliser les instruments présentés précédemment, il faut que les données empiriques que l'on possède, sur la structure rectangulaire, soient stationnaires. Gatrell (1979b) propose une façon ingénieuse de le tester. De façon globale, sa méthode va comme suit : si les données sont spatialement stationnaires, il faut que dans l'ensemble, les moyennes, variances et covariances échantillonnales soient invariantes par rapport à la localisation.

---

[1]. Gatrell (1977) et (1979b), sans s'assurer de la pertinence de la procédure, par analogie aux séries temporelles, considère que :

$$z(s,t) = \frac{R(s,t)}{\sqrt{(P-s)\ (Q-t)}} \sim N(0,\ 1)\ .$$

Donc, en termes pratiques, si on divise la région en quelques sous-régions et que l'on évalue ces quantités, en utilisant les mesures qu'on va présenter au chapitre II, on peut alors tester si les valeurs numériques de ces estimés semblent se répartir de façon aléatoire. Si non, il conseille de stationnariser en prenant une première différence spatiale [Martin (1974)].

2.3.2) *La spécification dans les deux modèles autorégressifs isotropes de Haining*

Haining (1978b), chapitre 3, veut savoir s'il est possible, à partir des autocorrélations empiriques R(s,t), de discriminer entre les modèles isotropes (52) et (53) [1] :

$$X_{i,j} = \alpha(X_{i-1,j} + X_{i,j-1}) + u_{i,j} \qquad i,j \in \mathbb{Z}^2$$

et

$$X_{i,j} = \rho(X_{i-1,j} + X_{i+1,j} + X_{i,j-1} + X_{i,j+1}) + u_{i,j} \qquad i,j \in \mathbb{Z}^2 .$$

Pour ce faire, Haining se base sur des résultats valables dans le cas de champs aléatoires continus. Pour bien comprendre, revenons aux modèles temporels unilatéral et bilatéral du début du chapitre :

(58) $$X_t = \theta X_{t-1} + u_t \qquad , \quad t \in \mathbb{Z}$$

et

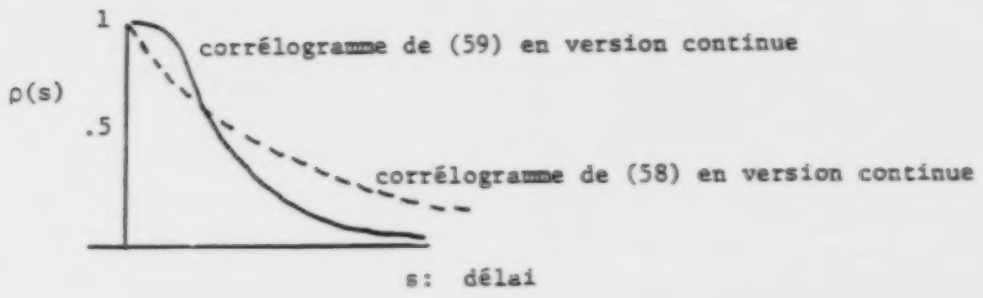
(59) $$X_t = \gamma(X_{t-1} + X_{t+1}) + u_t \qquad , \quad t \in \mathbb{Z}$$

---

[1]. On peut vérifier que les tableaux de corrélations théoriques de ces deux modèles pour $\alpha = 2\rho$ sont très semblables (voir Haining (1977), p. 126-27).

Le processus stationnaire continu formé à partir de la version continue de (58) possède une fonction d'autocorrélation théorique de type exponentiel (convexe) alors que le corrélogramme de la version continue de (59) possède à l'origine, comme illustré à la figure 6, une partie concave pour devenir convexe par la suite.

FIGURE 6



La différence dans l'allure théorique de ces deux versions permet au chercheur de discriminer entre les deux. Haining montre que la version continue de (52) produit une fonction d'autocorrélation exponentielle et la version continue de (53) une fonction concave à l'origine et convexe par la suite. Voir Haining (1977), pp. 115-116.

Pour un champ aléatoire discret (le cas qui nous concerne ici), Haining espère que le phénomène vaudra encore, c'est-à-dire que le modèle (53) par opposition au modèle (52) devrait posséder, dû à la concavité à l'origine, des autocorrélations théoriques qui respectent :

(60) $\quad 1 - \rho(1,0) < \rho(1,0) - \rho(2,0)$,

ou

(61) $\quad 1 - \rho(1,1) < \rho(1,1) - \rho(2,2)$,

c'est-à-dire que pour $t = 0$ en (60), les corrélations sur l'axe s devraient diminuer plus entre les délais 1 et 2 que les délais initiaux 0 et 1.

Après avoir calculé les tableaux d'autocorrélations théoriques de ces deux modèles pour différentes valeurs de α entre 0 et ,5 et ρ entre 0 et ,25, il constate que le phénomène remarqué dans les versions continues ne tient plus (en général) pour le cas discret. Les tableaux de corrélations théoriques formées à partir des modèles (52) et (53) produisent pour des valeurs $\alpha \approx 2\rho$, un comportement fortement similaire. D'après ses résultats, les formules (60) et (61) ne nous permettent généralement pas de choisir entre (52) et (53). Le problème devient évidemment plus grave lorsqu'on considère les corrélations empiriques. Comme solution, Haining propose de choisir le modèle qui maximise la vraisemblance lors de l'estimation des deux modèles.

Avant de terminer cette section, on aimerait formuler une remarque supplémentaire sur la spécification du modèle (53).

Remarque

Concernant l'évaluation des autocorrélations et autocovariances du modèle Q.A. (53), à cause de la formulation bilatérale (symétrique) de ce modèle l'erreur $u_{i,j}$ est corrélée avec $X_{i-1,j}$, $X_{i+1,j}$, $X_{i,j-1}$, $X_{i,j+1}$ $\forall_{i,j} \in \mathbb{Z}^2$.

Par exemple, pour $E(X_{i,j}) = 0$ et

$X_{i,j} = \rho(X_{i-1,j} + X_{i+1,j} + X_{i,j-1} + X_{i,j+1}) + u_{i,j}$, $i,j \in \mathbb{Z}^2$, on a :

$$(62) \quad \sigma(0,1) = E(X_{i,j} \cdot X_{i,j+1}) = \rho[E(X_{i-1,j} \cdot X_{i,j+1}) + E(X_{i+1,j} X_{i,j+1}) + E(X_{i,j-1} X_{i,j+1}) + E(X_{i,j+1} X_{i,j+1})] + \underbrace{E(u_{i,j} \cdot X_{i,j+1})}_{=0} .$$

Cette dernière remarque nous permet d'introduire dans la discussion les travaux de Gebizlioglu (1982) et (1983) qui, par une extension des outils méthodologiques de la série chronologique développe les propriétés théoriques des modèles AR, MA et ARMA spatiaux. Selon ses résultats, pour spécifier l'ordre d'un MA spatial, on examine l'ordre auquel les autocorrélations empiriques coupent. De la même façon, pour un modèle autorégressif pur spatial, il propose les autocorrélations partielles. De façon globale, bien que l'ensemble de ses résultats théoriques semblent vraisemblables, les hypothèses sur lesquelles il fonde son approche le sont moins.

En effet, selon son ensemble d'hypothèses, qu'il considère minimum, il pose que $E(u_{i,j} \cdot X_{i+s,j+t}) = 0$ si s et t ne sont pas nuls en même temps. Ce qui veut dire qu'à l'équation (62), il pose $E(u_{i,j} \cdot X_{i,j+1}) = 0$, ce qui est tout-à-fait incorrect sauf pour des modèles spatiaux unilatéraux comme le modèle en vague. Ce genre d'hypothèse n'est pas raisonnable, pour un modèle bilatéral. C'est toutefois ce qui lui permet de construire les équations de Yule-Walker sur lesquelles sont basées les autocorrélations partielles. Cette approche est à nos yeux injustifiée et c'est pourquoi on ne se contente ici que d'informer le lecteur sur la présence d'une telle approche dans la littérature.

3) Modèles autorégressifs spatiaux pour une configuration irrégulière

Au début du chapitre, on a fait état de la présence de deux tendances dans l'étude des processus spatiaux. C'est ici qu'on traite de la seconde. Elle origine en fait des travaux de Cliff et Ord entrepris vers la fin des années 1960. Bien que leurs recherches au début furent concentrées sur le développement de tests d'autocorrélation spatiale[1], ces efforts ont finalement conduit ces auteurs à proposer une classe de modèles ARMA spatiaux (SARMA), applicables à la modélisation de phénomènes qui se réalisent sur ces configurations géographiques irrégulières[2]. Pour eux, une configuration spatiale est habituellement irrégulière et il faut développer des instruments qui tiennent compte de ce fait [voir Cliff et Ord (1975a)].

---

1. Voir chapitre II.

2. Pour que les outils des deux premières sections soient applicables pour pour une configuration irrégulière, Haining (1978b) propose de former une approximation rectangulaire de la région.

Les modèles qu'ils proposent sont de type isotrope. Cette hypothèse est forte car, comme le souligne Gatrell (1977), lorsque le tableau des corrélations empiriques R(s,t) laisse voir des tendances directionnelles, il est important d'en tenir compte. Bien que l'isotropie soit facilement critiquable, le fait de postuler la stationnarité spatiale l'est presqu'autant. En effet, Granger (1969) trouve que pour des variables économiques, l'hypothèse de stationnarité spatiale est discutable. En effet, selon lui, il n'est pas raisonnable de croire que la relation entre les valeurs mesurées $x_{i,j}$ et $x_{i+s,j+t}$ est la même pour toutes paires de zones dont les positions relatives sont les mêmes.

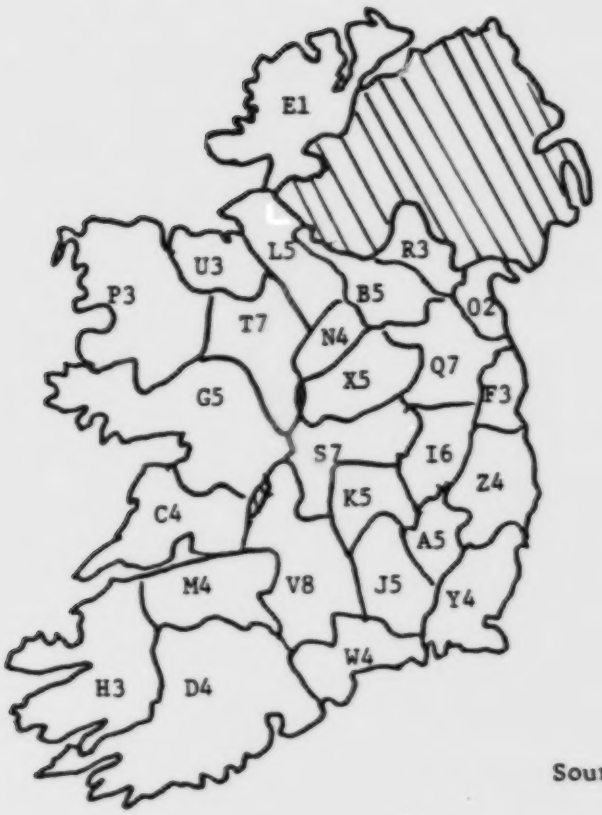
Malgré toutes les critiques qu'on peut faire, la classe SARMA développée par Cliff et Ord (1975c) s'avère extrêmement populaire en géographie comme en économie. D'ailleurs, au chapitre III, on pourra constater l'abondance d'applications qui utilisent la version autorégressive SAR(1). Pour notre part, nous nous intéresserons dans les prochains chapitres, au modèle SAR(1) appliqué aux résidus d'une régression.

Avant d'aller plus loin, présentons maintenant les concepts méthodologiques de cette approche.

3.1) Concepts et notation

Généralement, l'organisation géographique de zones forme une région dont la configuration est rarement rectangulaire. Par exemple, comme application empirique, on s'intéresse à la république d'Irlande. C'est ce que l'on présente à la figure 7, où la lettre identifie une des 26 zones et le

FIGURE 7



Source : Geary (1968)

chiffre, le nombre de voisins[1]. Prendre une approximation rectangulaire de cette région n'est pas une procédure à recommander ici. Le problème est d'autant plus sérieux que les zones ne sont pas toutes de même dimension.

Ici, on considère que l'on possède une observation x_i pour chaque zone i, i=1,...,N. Le vecteur x est considéré comme la réalisation du vecteur aléatoire $X = [X_1,...,X_N]$, (notre échantillon)[2]. Pour ce genre de problème, si on veut modéliser une forme autorégressive, il faut introduire le concept de contiguïté.

---

[1]. Voir appendice I pour de l'information sur les données de l'étude.

[2]. Comme c'est souvent le cas dans la littérature, le contexte étant toujours clair, à partir de maintenant, on ne fera plus de différence de notation entre l'échantillon X et la réalisation x. On emploiera toujours x.

i) Contiguïté physique d'ordre 1

Toute zone i, qui a une frontière commune avec une zone j, est dite contiguë physique d'ordre 1. Pour retenir cette information, on utilise la matrice de contiguïté.

Une matrice N×N de contiguité C est telle que :

$$(63) \qquad c_{ij} = \begin{cases} 1 & \text{si i et j sont contiguë d'ordre 1} \\ 0 & \text{sinon ,} \end{cases}$$

avec, évidemment $c_{ii} = 0 \ \forall i$. Si on veut parler de plusieurs ordres de contiguïté physique, Hordijk (1974) propose la définition suivante :

ii) Contiguïté physique d'ordre g

Pour $g > 1$, une zone i est contiguë d'ordre g, à une zone j ($\neq i$) si i est contiguë de premier ordre à une zone h, contiguë d'ordre g-1, à j, et si i n'est pas déjà contiguë à j d'un ordre inférieur à g.

Pour chaque ordre de contiguïté, on peut définir une matrice C(g) où :

$$(64) \qquad c_{ij}(g) = \begin{cases} 1 & \text{si i et j sont contiguës d'ordre g} \\ 0 & \text{sinon .} \end{cases}$$

Soit G, le nombre maximum d'ordre de contiguïté de la région, on peut vérifier qu'on aura toujours :

$$(65) \qquad I + C(1) + C(2) + \ldots + C(G) = ii'$$

où i est le vecteur unitaire N×1, $i = [1,1,\ldots,1]'$.

Pour une région subdivisée en un grand nombre de zones, la construction de ce genre de matrices peut rendre la tâche longue et ennuyeuse. C'est pourquoi Cliff et Ord (1981) proposent une approche exprimée en terme de distances. En effet, choisissons à priori G distances de références que l'on note :

$$(66) \qquad (a_1, a_2, \ldots, a_G) \quad ,$$

$$a_1 < a_2 < \ldots < a_G \quad \text{et} \quad a_o = 1 \ .$$

Alors C(g) devient la matrice de contiguïté d'ordre g où :

$$(67) \qquad c_{ij}(g) = \begin{cases} 1 & \text{si } a_{g-1} \leq d_{ij} < a_g \\ 0 & \text{sinon} \end{cases} ,$$

où d_{ij} représente la distance entre les centroïdes des zones i et j.

Pour notre part, comme on se limitera à la contiguïté d'ordre 1, on pose comme convention :

$$(68) \qquad C = C(1) \quad .$$

Le modèle Q.A. isotrope (53) qui génère le vecteur aléatoire x de dimensions PQ×1 :

$$(69) \qquad x_{i,j} = \rho(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}) + u_{i,j} \quad , \qquad \begin{matrix} i = 1,\ldots,P \\ j = 1,\ldots,Q \end{matrix}$$

$$u_{i,j} \sim BB(0, \sigma^2) \ ,$$

s'écrit pour une structure irrégulière qui possède N zones :

$$(70) \qquad x_i = \rho \sum_{j=1}^{N} c_{ij}\, x_j + u_i \quad , \qquad i = 1,\ldots,N,$$

où $c_{ii} = 0 \quad \forall i$ et $u_i \sim BB(0,\sigma^2)$.

De façon matricielle, le modèle (70) devient : [SAR(1)]

$$(71) \qquad \underset{(N\times 1)}{x} = \rho \underset{(N\times N)}{C} \underset{(N\times 1)}{x} + \underset{(N\times 1)}{u} \quad , \qquad u \sim BB(0,\ \sigma^2 I_N) \ .$$

Pour Bartels (1979), ce genre de modèle constitue une adaptation pour le contexte spatial des concepts de l'analyse des séries temporelles. En effet, comme hypothèse, Cliff et Ord (1981) considèrent que :

$$(72) \qquad \begin{aligned} E(x_i) &= \mu \quad , && \forall i \\ V(x_i) &= \sigma^2 \quad , && \forall i \\ E[x_i\, x_i(g)] &= \sigma(g) \quad , && \forall g \text{ et } \forall i \ . \end{aligned}$$

où $x_i(g) = \sum_{j=1}^{N} c_{ij}(g)\, x_j$ constitue l'ensemble de voisins d'ordre g de la zone i.

Ceci veut dire, par exemple pour g = 1, que le lien entre une zone i et l'ensemble de ses voisins d'ordre 1 est le même pour toute zone i. Cette hypothèse est moins raisonnable que dans le cas d'une structure rectangulaire comme le modèle (69)[1]. Malgré cela, de la même façon qu'un AR(1) est généralement postulé en séries temporelles, le SAR(1) (71) l'est pour des données spatiales.

---

[1]. Pour une formulation irrégulière, on peut vérifier facilement que les moments de x en (71) ne respectent pas (72). C'est pourquoi, plus loin, on préférera parler de SAR(1) stable plutôt que stationnaire.

Remarque

Comme on veut surtout porter l'emphase sur les problèmes d'estimation du SAR(1), on ne présentera pas ici les outils de spécification de la version générale SARMA(p,q). Au Chapitre 5 de Cliff et Ord (1981), on retrouve un ensemble d'outils inspirés de la série temporelle qui permettent de faire de la spécification. Par exemple, avec l'aide de corrélogrammes spatiaux, on peut identifier l'ordre d'un SMA(q). Grâce à un corrélogramme partiel spatial, on peut déterminer l'ordre d'un SAR(p). Pour plus d'informations, consultez cette référence.

Au chapitre III, on s'intéressera aux propriétés théoriques et empiriques de la version SAR(1) appliquée aux résidus d'un modèle de régressions, c'est-à-dire qu'on étudiera le modèle suivant :

$$Y = X\beta + \varepsilon \quad , \tag{73}$$

$$\varepsilon = \rho C \varepsilon + u \quad , \qquad u \sim N(0, \sigma^2 I) \; .$$

Pour Cliff et Ord, la matrice C est critiquable en ce sens qu'elle ne transporte pas d'informations sur la structure physico-économique de la région. Par exemple, à la figure 7, les zones ne sont pas toutes de mêmes dimensions. De plus, la distance entre les centroïdes des zones peut être un facteur qui affecte l'intensité de la contiguïté entre les zones. Plus les zones i et j sont éloignées, moins elles devraient être liées.

Ils proposent donc de remplacer dans les schémas (70), (71) et (73) les $c_{ij} = 1$ par des éléments $w_{ij} \neq 0$ où w_{ij} sont des fonctions fixées avant estimation. Par exemple, on pourrait avoir :

$$w_{ij} = \begin{cases} d_{ij}^{-\theta} & \text{si i et j sont contiguës d'ordre 1, où } \theta \text{ est fixé à 1 ou 2,} \\ 0 & \text{ailleurs.} \end{cases} \tag{74}$$

Il est évident que pour un ordre g de contiguïté, on aura comme matrice de contiguïté non booléenne :

$$W(g) \ . \tag{75}$$

3.2) Les matrices de contiguïté normalisées

Dans la littérature, pour donner une interprétation intuitive à ρ, les auteurs proposent habituellement de normaliser W comme suit :

$$m_{ij} = \frac{w_{ij}}{\sum_j w_{ij}} \ . \tag{76}$$

De cette façon, même si W est une matrice symétrique, la normalisation (76) produit une matrice $\underset{(N\times N)}{M}$ non symétrique telle que :

$$\sum_{j=1}^{N} m_{ij} = 1 \qquad \forall i \ . \tag{77}$$

Pour un problème où on possède un nombre important de zones, dans la formulation de l'algorithme d'estimation des modèles (71) et (73), il peut être intéressant de conserver la propriété de symétrie. Comme la normalisation, on le verra, est désirable, on propose dans notre approche la matrice de contiguïté normalisée symétrique, $\underset{\sim}{W}$, où :

$$(78) \qquad \underset{\sim}{w}_{ij} = \frac{w_{ij}}{\sum_{i=1}^{N} w_{ij} \cdot \sum_{j=1}^{N} w_{ij}} \quad .$$

On pourra constater que la matrice $\underset{\sim}{W}$ a les mêmes propriétés intéressantes que la matrice M et est tout comme W symétrique. Dans les cas que l'on étudiera, du point de vue interprétation économique de la formulation, $\underset{\sim}{W}$ sera supérieure à M ou W.

Remarque

Un schéma autorégressif SAR(1) s'écrit :

$$(79) \qquad y = \rho W y + u \ .$$

Lorsqu'on remplace m fois le vecteur y à droite de l'expression, on obtient :

$$(80) \qquad y = \rho^m W^m y + \sum_{i=0}^{m-1} \rho^i W^i u$$

Ainsi, lorsque $m \to \infty$, y est exprimable comme une somme sur le vecteur stationnaire u. Pour que cette somme soit convergente, il faut que :

$$(81) \qquad |\rho| < \frac{1}{|\lambda_{max}|}$$

où $|\lambda_{max}|$ est la valeur absolue de la racine caractéristique dominante de W (voir la section 4 du chapitre III).

Ainsi, lorsque la condition (81) est respectée, le modèle $y = \sum_{i=0}^{\infty} \rho^i W^i u$ est valide au sens de la convergence en moyenne quadratique. D'où on dira que le processus autorégressif SAR(1) (70) est stable

lorsque (81) est respecté. Comme mentionné à la note de bas de page précédente, pour un modèle spatial irrégulier, les moments ne respectent pas (72). On préfère donc parler de stabilité plutôt que de stationnarité.

CHAPITRE II

TESTS D'AUTOCORRÉLATION SPATIALE

Considérons une région avec N zones, où pour chacune d'elles on possède une variable aléatoire X_i dont la réalisation est x_i. La loi de distribution jointe de l'échantillon $\underset{(N\times 1)}{X}$ s'écrit :

$$(1) \qquad P(X_1 < x_1,\ X_2 < x_2,\ \ldots,\ X_N < x_N) = F(x_1,\ x_2, \ldots, x_N) \ .$$

Lorsque $P(X_1 < x_1,\ X_2 < x_2,\ \ldots,\ X_N < x_N)$ s'exprime comme $\prod_{i=1}^{N} P(X_i < x_i)$, notre échantillon $\underset{(N\times 1)}{X}$ devient un échantillon aléatoire, c'est-à-dire que chaque X_i peut alors être considéré comme issu d'un tirage aléatoire indépendant. Comme l'indépendance implique l'absence de corrélation, on a :

Définition 1.

Soit $\underset{(N\times 1)}{X}$ le vecteur aléatoire dont la réalisation s'écrit $\underset{(N\times 1)}{x}$. Lorsque

$$(2) \qquad P(X_1 < x_1,\ X_2 < x_2,\ \ldots,\ X_N < x_N) = \prod_{i=1}^{N} P(X_i < x_i) \ ,$$

il y a alors absence de corrélation spatiale. Comme on traite ici d'une même série $\{X_i\}$, il devient plus pertinent de parler d'autocorrélation spatiale.

Depuis Moran (1948), de nombreuses mesures ont été développées afin de tester, pour des données spatiales qualitatives ou continues, l'expression (2). Pour une présentation extensive de toutes les mesures d'autocorrélation spatiale existantes, consulter les chapitres 1 et 2 de Cliff et Ord (1981).

En prenant comme source principale cette dernière référence, on présente à la section 1) les mesures qui visent à tester si N observations x_i continues sont spatialement correlées. Pour les mesures qui risquent le plus d'intéresser les économistes, par une approche qui, contrairement à celle de Cliff et Ord, est exclusivement matricielle, on développera les moments de distribution de ces statistiques. Comme ils l'ont fait, on pourra, par la suite, vérifier que sous des conditions généralement respectées, ces mesures ont une loi de distribution asymptotique normale. De la même façon, à la section 2), on étudiera les mesures qui permettent de tester l'autocorrélation spatiale dans les résidus d'une régression.

1) Mesures d'autocorrélation spatiale pour des données continues

Lorsque les x_i de valeurs similaires tendent à se grouper spatialement, il y a de l'autocorrélation spatiale positive. Avec de l'autocorrélation spatiale négative, les x_i semblables cherchent plutôt à s'éparpiller. On a absence d'autocorrélation spatiale lorsque les x_i se répartissent de façon aléatoire. Par exemple, à la figure 1, en se limitant aux signes des x_i, on fait face à un problème d'autocorrélation positive. Pour plus de détails sur cette figure, se reporter au chapitre V.

FIGURE 1

Pour des données continues, il existe 4 mesures d'autocorrélation spatiale. Elles ont été proposées par Moran (1950), Geary (1968), Dacey (1965) et Cliff et Ord (1969). On veut, dans un premier temps, exposer les caractéristiques de chacune d'elles.

1.1) Mesure de Moran

$$(3) \qquad I_x^M = \frac{N}{2A} \frac{\sum_{(2)} c_{ij} z_i z_j}{\sum_{i=1}^{N} z_i^2}$$

où N = le nombre de région, $z_i = x_i - \bar{x}$, $2A = i' C i$

où i est le vecteur unitaire d'ordre $N \times 1$ et C la matrice de contiguïté physique d'ordre 1, où, on le rappelle $c_{ij} = \begin{cases} 1 & \text{si } i \text{ et } j \text{ sont contigües} \\ 0 & \text{sinon,} \end{cases}$

et de plus $\sum_{(2)} = \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N}$.

Selon les conventions que l'on établit ici, I_x^M représente la mesure d'autocorrélation spatiale de l'auteur Moran (M) applicable sur les variables continues (x). Pour Geary, Dacey et Cliff et Ord, on a respectivement I_x^G, I_x^D, I_x^{CO}.

La région est spatialement correlée, de façon positive, lorsque $I_x^M > 0$. Pour $I_x^M < 0$, il y a de la corrélation spatiale négative. Comme mesure similaire à la statistique de Durbin-Watson utilisée en séries chronologiques, on présente maintenant la mesure de Geary.

1.2) Mesure de Geary

$$(4) \qquad I_x^G = \frac{(N-1)}{4A} \cdot \frac{\sum_{(2)} c_{ij} (z_i - z_j)^2}{\sum_{i=1}^{N} z_i^2} \quad .$$

Cliff et Ord (1981) mentionnent que, sous l'hypothèse nulle d'absence de corrélation spatiale, I_x^G et I_x^M ont toutes deux une loi de distribution limite normale. En considérant que cette approximation est valable pour un échantillon de taille raisonnable ($N > 50$), ces auteurs proposent de tester la corrélation spatiale, en comparant la mesure standardisée $Z(I_x)$ à une $N(0,1)$ à $(1-\alpha)\%$. Par exemple, pour I_x^M, on a :

$$(5) \qquad Z(I_x^M) = \frac{I_x^M - E(I_x^M)}{\sigma(I_x^M)} \sim N(0,1) \quad .$$

Pour Dacey (1965), une critique générale s'applique à ces deux mesures. Elles sont sujettes à l'invariance topologique. Explicitons premièrement le concept d'invariance topologique.

1.3) Invariance topologique

Calculons pour la région de la figure 1, la valeur numérique de I_x^M. Disons que I_x^M est alors égal à un nombre quelconque Y_1. Par la suite, changeons

la configuration physique de la région, tout en maintenant la même structure de contiguïté entre les zones[1]. La valeur calculée de I_x^M pour cette nouvelle région, on peut le vérifier, demeure alors quand même Y_1. C'est ce que Dacey (1965) appelle l'invariance topologique.

Cet auteur propose alors une généralisation de I_x^M qui tient compte de la taille des zones et de la proportion de périmètre commun entre elles. Toutefois, cette mesure I_x^D est telle qu'il est, pour l'instant, difficile d'évaluer les moments de distribution. C'est pourquoi Cliff et Ord proposent une statistique qui, à la fois, possède des moments faciles à évaluer et qui conserve l'intérêt intuitif de la mesure de Dacey.

1.4) Mesure de Cliff et Ord

$$(6) \qquad I_x^{CO} = \frac{N}{S_0} \frac{\sum_{(2)} w_{ij} z_i z_j}{\sum_{i=1}^{N} z_i^2}$$

où $S_0 = i' W i$, et W est la matrice de contiguïté non booléenne présentée au chapitre I, dont les éléments w_{ij} sont nuls si les zones i et j sont non contiguës. La similitude entre (6) et (3) explique pourquoi certains appellent la mesure I_x^{CO} mesure de Moran généralisée.

Comme on l'a déjà dit, les w_{ij} sont des fonctions fixées à priori qui permettent de décrire la structure économico-géographique de la région. Comme les utilisateurs préfèrent généralement I_x^{CO} et I_x^M, on se limitera ici à l'étude de seulement ces deux mesures. Mais avant, formulons de façon matricielle les équations (3) et (6).

---

[1]. C'est-à-dire qu'il faut modifier la forme et la taille des zones tout en conservant pour chaque zone i le même ensemble de voisins du premier ordre.

1.5) Formulation matricielle de I_x^M et I_x^{CO}

Grâce aux matrices de contiguïté d'ordre 1 W et C, on a :

(7) $$I_x^M = \frac{N}{2A} \frac{z' C z}{z' z} \quad ,$$

où $2A = i' C i$, et

(8) $$I_x^{CO} = \frac{N}{S_0} \frac{z' W z}{z' z} \quad ,$$

où C et W sont des matrices symétriques.

Remarques

1. Dans une forme quadratique $z' W z$ où W n'est pas symétrique, il existe toujours un équivalent symétrique :

   (9) $$W_s = \frac{1}{2} (W + W') \quad ,$$

   tel que $z' W_s z$ est égal à $z' W z$. Il n'est alors pas contraignant de poser la symétrie pour la matrice W de l'équation (8).

2. lorsqu'on considère, à l'équation (8) l'équivalent symétrique de la matrice de contiguïté normalisée M, "$W = \frac{1}{2} (M + M')$", puisque l'on a $M \cdot i = i$ et $i' M' = i'$, alors :

   (10) $$S_0 = i' W i \equiv N.$$

   Dans ce cas la mesure I_x^{CO} devient[1] :

   (11) $$I_x^{CO} = \frac{z' W z}{z' z} \quad .$$

---

[1]. Il est évident que cette normalisation sur C conduit à :

$$I_x^M = \frac{z' C z}{z' z} \quad .$$

1.6) Théorie de distribution de I_x^{CO} et I_x^M

Comme I_x^M est un cas particulier de I_x^{CO} où $C = W$, on étudiera ici les propriétés statistiques de I_x^{CO} seulement.

1.6.1) *Moments de la loi de distribution d'échantillonnage de* I_x^{CO}

On considérera ici, sous l'hypothèse nulle d'absence de corrélation spatiale, que $X \sim N(0, \sigma^2 I)$. Comme convention, on ne fera pas de différence de notation entre l'échantillon aléatoire $\underset{(N\times 1)}{X}$ et l'échantillon observé $\underset{(N\times 1)}{x}$. Alors, on a :

$$H_0 : x \sim N(0, \sigma^2 I_N) \quad . \tag{12}$$

Ainsi, comme $\underset{(N\times 1)}{z} = x - i\bar{x}$, on obtient :

$$z = [I_N - \frac{1}{N} ii'] \; . = Ax \; , \tag{13}$$

où $A \cdot A = A = A'$, et $tr(A) = N - 1 = \rho(A)$, d'où

$$\underset{(N\times 1)}{z} \sim N(0, \sigma^2 \underset{(N\times N)}{A}) \; . \tag{14}$$

Afin d'évaluer les moments de distribution de la mesure I_x^{CO}

$$I_x^{CO} = \frac{N}{S_0} \frac{z'Wz}{z'z} = \frac{N}{S_0} \cdot \frac{x'AWAx}{x'Ax} \quad , \tag{15}$$

où $A' = A$, nous avons besoin du prochain théorème.

1.6.1.1) Preuve de l'applicabilité du théorème Pitman-Koopmans

Théorème Pitman-Koopmans

Soit Y un vecteur (N×1) tel que $Y \sim N(0, I_N)$. Toute statistique h(Y), invariante dans le changement d'unité (scale free) "$h(\lambda Y) = h(Y)$" a une loi de distribution indépendante de la forme quadratique $Q = Y'Y$.

Preuve : Appendice II, section 1)

Lorsque I_x^{CO} est exprimé comme à l'équation (15), ce théorème ne semble pas directement applicable. Toutefois, comme A et AWA sont des matrices symétriques telles que $A \cdot (AWA) = (AWA) \cdot A$, il existe alors une matrice orthogonale H qui diagonalise simultanément A et AWA[1], c'est-à-dire qu'il existe une matrice orthogonale H telle que[2] :

$$(16) \qquad H' A H = D \qquad \text{et} \qquad H' AWA H = \Theta \ ,$$

où évidemment D et Θ sont les matrices diagonales des valeurs propres de A et AWA respectivement. Comme I_x^{CO} est une mesure invariante par rapport à l'échelle, on peut définir que $x \sim N(0, I_N)$. Grâce à la transformation orthogonale $y = H' x$ qui implique $y \sim N(0, I_N)$, la mesure I_x^{CO} peut s'écrire de façon équivalente :

$$(17) \qquad I_x^{CO} = \frac{N}{S_0} \cdot \frac{x'AWA\,x}{x'A\,x} = \frac{N}{S_0} \frac{y'\Theta\,y}{y'D\,y} \ .$$

---

1. Voir Dhrymes (1978), proposition 53, p. 481.

2. Pour le développement algébrique de H, voir Dhrymes (1978), p. 146-148, ou Durbin-Watson (1950), p. 412-413.

Comme A est une matrice idempotente de rang (N-1), on peut écrire :

$D = \begin{bmatrix} I_{N-1} & 0_{(N-1)\times 1} \\ 0_{1\times(N-1)} & 0_{(1\times 1)} \end{bmatrix}$. Ce dernier résultat permet d'établir que Θ aura au plus N-1 éléments ν_i non nuls[1], avec : $\Theta = \begin{bmatrix} \nu_1 & & 0 \\ & \ddots & \\ 0 & & \nu_{N-1} \end{bmatrix}$. I_x^{CO} s'écrit alors, grâce à ces résultats :

$$(17.1) \qquad I_x^{CO} = \frac{N}{S_0} \frac{\sum_{i=1}^{N-1} \nu_i y_i^2}{\sum_{i=1}^{N-1} y_i^2} \quad .$$

Le théorème Pitman-Koopmans peut donc s'appliquer sur (15).

A cause de cette indépendance entre I_x^{CO} et $Q = z'z$, on peut écrire :

$$(18) \qquad E(I_x^{CO}) = E(I_x^{CO}) \cdot \frac{E(Q)}{E(Q)} = \frac{E(I_x^{CO} \cdot Q)}{E(Q)} \quad ,$$

et donc, on a de façon générale :

$$(19) \qquad E[(I_x^{CO})^p] = (\frac{N}{S_0})^p \cdot \frac{E[(z'Wz)^p]}{E[(z'z)^p]} \quad , \qquad p = 1,2,\ldots \quad .$$

1.6.1.2) Dérivation des deux premiers moments de distribution de I_x^{CO}

I_x^{CO}, par (15), s'écrit :

$$(20) \qquad I_x^{CO} = \frac{N}{S_0} \cdot \frac{x'AWAx}{x'Ax} \quad ,$$

où $x \sim N(0, \sigma^2 I_N)$.

---

[1]. Voir Durbin et Watson (1950), p. 412-13.

Par les théorèmes 1 et 2 de l'Appendice II.2), sur les moments de formes quadratiques formées à partir de variables normales indépendantes, on a :

$$E(x' A x) = \sigma^2 \text{ tr } (A) = \sigma^2 (N-1)$$

$$E(x' A W A x) = \sigma^2 \text{ tr } (WA).$$

Comme $A = [I - \frac{1}{N} ii']$ et $tr(W) = 0$, on obtient :

$$(21) \qquad E(I_x^{CO}) = \frac{-N}{S_0} \frac{tr\ (W\ ii')}{N\ (N-1)} = - \frac{1}{(N-1)} \quad ,$$

où $S_0 = i' W i$.

Comme :

$$E[(x' A x)^2] = 2\ \sigma^4\ tr(A\ I\ A\ I) + \sigma^4\ tr^2\ (A\ I) \ ,$$

$$= 2\ \sigma^4\ (N-1) + \sigma^4\ (N-1)^2 \quad , \text{ car } tr(A) = N-1 \ ,$$

$$= \sigma^4\ (N+1)\ (N-1) = \sigma^4\ (N^2 - 1) \quad ,$$

et

$$E[(x' A W A x)^2] = \sigma^4\ [2\ tr\ (A\ W\ A\ A\ W\ A) + tr^2\ (A\ W\ A)] \ ,$$

alors,

$$(22) \qquad E[(I_x^{CO})^2] = \frac{N^2}{S_0^2\ (N^2-1)}\ [2\ tr\ (W\ A\ W\ A) + tr^2\ (W\ A)] \quad .$$

Pour avoir $V(I_x^{CO})$, on calcule, grâce à (21) et (22) :

$$(23) \qquad V(I_x^{CO}) = E[(I_x^{CO})^2] - E^2(I_x^{CO}) \quad .$$

On peut facilement vérifier que l'expression (22) est équivalente à l'expression (1.38) de Cliff et Ord (1981). Quiconque n'est pas intéressé à ce développement peut sauter la prochaine section.

1.6.2) Equivalence entre (22) et (1.38) de Cliff et Ord (1981)

i) notation de Cliff et Ord

$$1) \quad S_0 = \sum_{(2)} w_{ij} = i' W i \quad ,$$

$$2) \quad S_1 = \frac{1}{2} \sum_{(2)} (w_{ij} + w_{ji})^2 \quad ,$$

$$3) \quad S_2 = \sum_{i=1}^{N} (w_{i.} + w_{.i})^2$$

où $w_{i.} = \sum_{j=1}^{N} w_{ij}$ et $w_{.i} = \sum_{j=1}^{N} w_{ji}$.

Comme expression de $E[(I_x^{CO})^2]$, ces auteurs obtiennent :

$$(1.38) \quad E[(I_x^{CO})^2] = \frac{N^2 S_1 - N S_2 + 3 S_0^2}{S_0^2 (N^2 - 1)} \quad .$$

Comme :

a) $tr(W^2) = tr(W W) = \sum_{i \neq j} \sum w_{ij}^2$,

car $w_{ii} = 0 \quad \forall i$. Comme par (9) W peut toujours être symétrique dans I_x^{CO}, alors :

$$S_1 = 2 \sum_{(2)} w_{ij}^2 ,$$

$$\text{d'où :} \quad ① \quad \text{tr}\,(W^2) = \frac{S_1}{2} \quad .$$

De la même façon :

$$\text{b)} \quad i' W W i = \sum_{i=1}^{N} w_{i.}^2 = \frac{S_2}{4} ,$$

$$\text{d'où :} \quad ② \quad i' W W i = \frac{S_2}{4} \quad .$$

Comme $A = [I - \frac{1}{N} ii']$, l'expression (22) devient :

$$E[(I_x^{CO})^2] = \frac{N^2}{S_0^2(N^2-1)} \cdot \left(2\ \text{tr}[W \cdot (W - \frac{ii'}{N} W)\ (I - \frac{ii'}{N})] + (0 - \frac{i'Wi}{N})^2 \right) ,$$

$$= \frac{N^2}{S_0^2(N^2-1)} [2\ \text{tr}(W^2) - \frac{4}{N} i' W W i + 2 \frac{S_0^2}{N^2} + \frac{S_0^2}{N^2}] ,$$

$$= \frac{N^2}{S_0^2(N^2-1)} [2\ \text{tr}(W^2) + 3 \frac{S_0^2}{N^2} - \frac{4}{N} i' W W i] .$$

Par ① et ②, on a finalement :

$$= \frac{N^2}{S_0^2(N^2-1)} [S_1 + 3 \frac{S_0^2}{N^2} - \frac{S_2}{N}] .$$

1.6.3) *Théorie de distribution asymptotique de* I_x^{CO}

On vérifiera maintenant que pour un vecteur z distribué normalement, la statistique I_x^{CO}, sous des conditions habituellement respectées, approche la normalité, lorsque $N \to \infty$.

Pour simplifier, on considérera la statistique I_x^{CO} présentée à l'équation (11) où $i'Wi = S_0 = N$. On a donc :

$$I_x^{CO} = \frac{z'Wz}{z'z} \quad . \tag{24}$$

Posons de plus $M = [I - \frac{1}{N} ii'] = A$. On a donc :

$$I_x^{CO} = \frac{x' M W M x}{x' M x} = \frac{x' T x}{x' M x} \quad , \tag{25}$$

où $T = MWM$ et $x \sim N(0, I)$, avec $\sigma^2 = 1$.

Afin de montrer que (25) a une distribution limite normale, Cliff et Ord (1981) proposent d'utiliser le théorème suivant, lequel est équivalent au théorème 1 de l'Appendice III. Sous des conditions d'existence des moments, on a :

THEOREME 1

Soit F(y) une distribution uniquement caractérisée par ses moments μ_k, et soit $\{F_N(y)\}$ $(N = 1,2,...)$ la séquence des distributions correspondant aux moments μ_{Nk}. Alors, si $\mu_{Nk} \to \mu_k$ lorsque $N \to \infty$, pour $k = 1,2,...$, on en déduit alors que $F_N(y) \to F(y)$ en tout point de continuité de F(y).

Pour notre cas, $F_N(y)$ correspond à la loi de distribution de notre statistique, pour un N donné, et F(y) la distribution normale, laquelle on le sait est uniquement caractérisée par ses moments μ_k. Si les moments μ_{Nk} de $F_N(y)$ tendent vers ceux de la normale F(y), alors $F_N(y)$ est une distribution qui approche la normalité lorsque $N \to \infty$.

1.6.3.1] *Digression statistique*

Avant de poursuivre, on veut établir premièrement les propriétés générales qui caractérisent la normale F(y).

A.1] *Fonction génératrice de moments (f.g.m)*

Sous des conditions d'existence de la f.g.m., si on peut trouver la f.g.m. d'une variable aléatoire Y, alors théoriquement on peut trouver la loi de distribution de cette variable, car pour une f.g.m. donnée, il n'existe qu'une seule fonction de distribution correspondante[1]. On a :

$$m_Y(t) = E(e^{tY}) = \int_{-\infty}^{\infty} e^{ty} f_Y(y)\, dy \quad . \tag{26}$$

Comme son nom l'indique, la f.g.m. permet de trouver les moments. En effet, en applicant la règle de Leibnitz sur la dérivation des intégrales,

$$\frac{\partial m_Y(t)}{\partial t^r} = \int_{-\infty}^{\infty} y^r e^{ty} f_Y(y)\, dy \quad ,$$

$$\left.\frac{\partial m_Y(t)}{\partial t^r}\right|_{t=0} = \int_{-\infty}^{\infty} y^r f_Y(y)\, dy = E(Y^r) = \mu'_r \quad . \tag{27}$$

---

[1]. Mood, Graybill et Boes (1974), p. 80.

où μ'_r représente le moment d'ordre r par rapport à l'origine zéro. D'une autre façon[1], on peut écrire pour (26) :

$$(28) \qquad m_Y(t) = \sum_{j=0}^{\infty} \mu'_j \cdot \frac{t^j}{j!} \quad .$$

Les moments se trouvent par la formule :

$$(29) \qquad \left.\frac{\partial m_Y(t)}{\partial t^r}\right|_{t=0} = \mu'_r \quad ,$$

où en fait μ'_j est le coefficient attaché à $\frac{t^j}{j!}$ dans le développement (28).

De la même façon, on trouve les moments centrés par rapport à la moyenne $E(Y) = \mu'_1$ par :

$$(30) \qquad m_{Y-\mu'_1}(t) = \sum_{j=0}^{\infty} \mu_j \frac{t^j}{j!} \quad ,$$

où

$$(31) \qquad \mu_r = \int_{-\infty}^{\infty} (y - \mu'_1)^r f_Y(y)\, dy \quad .$$

---

1. Par (26) et le développement en série de e^{ty},

$$m_Y(t) = \int_{-\infty}^{\infty} \left(\sum_{j=0}^{\infty} \frac{(ty)^j}{j!} \right) \cdot f_Y(y)\, dy \quad ,$$

$$= \int_{-\infty}^{\infty} \left[1 + ty + \frac{(ty)^2}{2!} \ldots\right] f_Y(y)\, dy \quad ,$$

$$= \int_{-\infty}^{\infty} f_y(y)\, dy + \int_{-\infty}^{\infty} ty\, f_Y(y)\, dy + \ldots \quad ,$$

$$= \left(1 + \mu'_1 t + \frac{\mu'_2 t^2}{2!} + \ldots\right) \quad .$$

A.2] *Appliqué à la normale*

La f.g.m. de la normale pour $Y \sim N(\mu, \sigma^2)$ est :

$$(32) \quad m_Y(t) = e^{(\mu t + \frac{\sigma^2 t^2}{2})} ,$$

ce que l'on note $e\{\mu t + \frac{\sigma^2 t^2}{2}\}$. De plus,

$$(33) \quad m_{Y-\mu}(t) = e\{\frac{\sigma^2 t^2}{2}\}$$

pour $Y-\mu \sim N(0, \sigma^2)$.

Si on fait le développement de (33) :

$$e\{\frac{\sigma^2 t^2}{2}\} = \sum_{j=0}^{\infty} \frac{1}{j!} \left[\frac{\sigma^2 t^2}{2}\right]^j = 1 + \frac{\sigma^2 t^2}{2 \cdot 1!} + \frac{\sigma^4 t^4}{2^2 2!} + \frac{\sigma^6 t^6}{2^3 3!} \cdots ,$$

on a :

$$(34) \quad m_{Y-\mu}(t) = 1 + \frac{\sigma^2 2!}{2 \cdot 1!} \cdot \frac{t^2}{2!} + \frac{\sigma^4 4!}{2^2 2!} \cdot \frac{t^4}{4!} \cdots .$$

Ainsi, par (30), les moments centrés par rapport à la moyenne d'une normale s'écrivent :

$$(35) \quad \begin{cases} \mu_{2k} = \dfrac{\sigma^{2k} \cdot (2k)!}{2^k \cdot k!} , & \text{pour } k = 1,2,\ldots , \\ \mu_{2k+1} = 0 , & \text{pour } k = 0,1,2,\ldots . \end{cases}$$

On définit le coefficient d'asymétrie (skewness) par :

$$(36) \quad \gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E(Y-\mu)^3}{[\sigma^2]^{3/2}} ,$$

et le coefficient d'applatissement (kurtosis) par :

$$(37) \qquad \frac{\mu_4}{\sigma^4} = \frac{E(Y-\mu)^4}{[\sigma^2]^2} \quad .$$

Rappelons que la normale est telle que $\gamma_1 = 0$ et $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = 0$.

B.1) *Fonction génératrice de cumulants* (f.g.c.)

Entre les cumulants K_j et les moments non centrés μ'_r, il existe la relation suivante[1] :

$$(38) \qquad e^{\left\{ \sum_{j=1}^{\infty} K_j \frac{t^j}{j!} \right\}} = m_Y(t) \quad ,$$

ou encore, en prenant le logarithme naturel de (38) :

$$(39) \qquad \sum_{j=1}^{\infty} K_j \cdot \frac{t^j}{j!} = \ln m_Y(t) \quad .$$

Alors, de la même façon que μ'_r est le coefficient de $\frac{t^r}{r!}$ dans (28), K_j est le coefficient de $\frac{t^j}{j!}$ dans (39). D'où $\ln m_Y(t)$ est la (f.g.c.) de Y. Alors par (38) et (39), Kendall et Stuart[2] établissent les relations suivantes entre moments et cumulants.

---

1. Kendall et Stuart (1963), p. 67. Ces auteurs soulignent qu'au point de vue théorique, les cumulants ont des propriétés plus intéressantes que les moments. On n'entrera toutefois pas dans ces considérations.

2. Kendall et Stuart (1963), p. 70-71.

Moments non centrés

$$\mu'_1 = K_1 \quad , \quad K_1 = \mu'_1 \quad ,$$

$$\mu'_2 = K_2 + K_1^2 \quad , \quad K_2 = \mu'_2 - \mu'^2_1 \quad , \tag{40}$$

$$\mu'_3 = K_3 + 3\,K_2K_1 + K_1^3 \quad , \quad K_3 = \mu'_3 - 3\,\mu'_2\,\mu'_1 + 2\,\mu'^3_1 \quad ,$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

Moments centrés

$$K_2 = \mu_2$$

$$K_3 = \mu_3 \tag{41}$$

$$K_4 = \mu_4 - 3\,\mu_2^2$$

$$\vdots \qquad \vdots \qquad \vdots \qquad .$$

B.2) *Appliqué à la normale centrée (non réduite)*

Soit

$$Y \sim N(0, \sigma^2) \quad . \tag{42}$$

Alors, bien que ce soit inutile de l'écrire, on a :

$$f_Y(y) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-y^2}{2\sigma^2}\right\} \quad . \tag{43}$$

Par (33) et (39) :

$$(44) \qquad m_Y(t) = e\{\frac{\sigma^2 t^2}{2}\}$$

et donc $\ln m_Y(t) = \sigma^2 \cdot \frac{t^2}{2} = K_2 \frac{t^2}{2}$.

Ces auteurs obtiennent donc que pour la normale :

$$(45) \qquad \begin{cases} K_2 = \sigma^2 \\ K_r = 0 \qquad\qquad r \geq 3 \end{cases} .$$

Par (41), on obtient aussi :

$$(46) \qquad \gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{K_3}{K_2^{3/2}}$$

$$(47) \qquad \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{K_4}{K_2^2} \quad .$$

Afin de démontrer la normalité asymptotique de I_x^{CO}, Cliff et Ord (1981) exploitent les expressions (46) et (47). En effet, on sait que la normale est telle que $K_r = 0$ pour $r \geq 3$.

1.6.3.2) *Normalité asymptotique de* I_x^{CO}

Grâce aux résultats établis à la section 1.6.1.1), on sait que :

$$(48) \qquad E(I_x^{CO})^p = \frac{E(x'T\,x)^p}{E(x'\,M\,x)^p} = \frac{E(y'\Theta\,y)^p}{E(y'\,D\,y)^p} \quad , \qquad p = 1,2,\ldots$$

où, d'après la notation établie aux équations (24) et (25), Θ est la matrice diagonale des racines caractéristiques de $T = M\,W\,M$ et D celle de M. De plus, $x \sim N(0, I_N)$ et $y = H'x \sim N(0, I_N)$.

Rappelons finalement que H est la matrice orthogonale qui diagonalise simultanément T et M. Donc $H'H = HH' = I_N$.

On sait que :

$$D = \begin{bmatrix} I_{N-1} & 0 \\ 0 & 0 \end{bmatrix}, \tag{49}$$

où 0 représente une matrice d'éléments zéro avec ordre approprié. On l'a déjà dit, pour une matrice W (N×N) symétrique quelconque, la matrice diagonale Θ aura au plus N-1 éléments υ_i non nuls[1]. A cause de (49), on a :

$$\Theta = \begin{bmatrix} \upsilon_1 & & 0 & 0 \\ & \ddots & & \\ & & \upsilon_{N-1} & \\ 0 & & 0 & 0 \end{bmatrix}. \tag{50}$$

Comme vu en (17.1), (49) et (50) permettent d'écrire :

$$I_x^{CO} = \frac{\sum_{i=1}^{N-1} \upsilon_i y_i^2}{\sum_{i=1}^{N-1} y_i^2}, \tag{51}$$

et

$$E(I_x^{CO})^p = \frac{E(\sum_{i=1}^{N-1} \upsilon_i y_i^2)^p}{E(\sum_{i=1}^{N-1} y_i^2)^p}. \qquad p = 1,2,... \tag{52}$$

---

1. Voir Durbin et Watson (1950), p. 412-413. Pour une matrice M de rang (N-k), on aura au maximum (N-k) éléments υ_i non nuls.

Pour simplifier la notation, écrivons $u' = \sum_{i=1}^{N-1} \upsilon_i \, y_i^2$ et $v' = \sum_{i=1}^{N-1} y_i^2$.

A cause de la décomposition (48) ou (52), on peut étudier la loi de distribution de I_x^{CO}, en évaluant indépendamment les moments de distribution de u' et v' [1].

Numérateur

Pour obtenir les moments de distribution de u', il s'agit de noter que u' est formé par une somme de variables aléatoires indépendantes $\upsilon_i \cdot y_i^2$ où $y_i^2 \sim \chi_1^2$. D'après Durbin et Watson (1950), le s ième cumulant de u' est alors la somme des (N-1) cumulants d'ordre s de $\nu_i \, y_i^2$, où i = 1,...,N-1. Ainsi, comme[2] :

$$(53) \qquad K_s \, (\upsilon_i \, y_i^2) = 2^{s-1} \, (s-1) \, ! \, \upsilon_i^s \quad ,$$

on a :

$$(53.1) \qquad K_s(u') = 2^{s-1} \, (s-1) \, ! \, \sum_{i=1}^{N-1} \upsilon_i^s \quad ,$$

et donc, en particulier, grâce à (40) et (41) :

$$(53.2) \qquad \mu_1'(u') = K_1(u') = \sum_{i=1}^{N-1} \upsilon_i \quad \text{et} \quad \mu_2(u') = K_2(u') = 2 \sum_{i=1}^{N-1} \upsilon_i^2 \quad .$$

---

[1]. Les prochains résultats sont fortement inspirés de Durbin et Watson (1950), p. 419-420.

[2]. Voir Durbin et Watson (1950), p. 419.

Dénominateur

Les moments non centrés du dénominateur v' sont simplement ceux d'une χ^2 avec N-1 degrés de liberté :

$$
\begin{aligned}
&E(v') = (N-1)\\
&E(v')^2 = (N-1)\ (N+1)\\
&E(v')^3 = (N-1)\ (N+1)\ (N+3)\\
&E(v')^4 = (N-1)\ (N+1)\ (N+3)\ (N+5) \quad , \text{ etc.}
\end{aligned}
\tag{54}
$$

Comme en B.1), on a établi les propriétés des cumulants d'une normale centrée non réduite, on aimerait ici obtenir les moments de la variable aléatoire centrée $r = I_x^{CO} - E(I_x^{CO})$. Par (53) et (54) :

$$E(I_x^{CO}) = K_1(I_x^{CO}) = \frac{\sum_{i=1}^{N-1} \upsilon_i}{N-1} = \bar{\upsilon} \quad ,$$

$$\text{d'où} \quad r = I_x^{CO} - E(I_x^{CO}) = \frac{\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})\ y_i^2}{\sum_{i=1}^{N-1} y_i^2} = \frac{u}{v'} \quad .$$

où évidemment on pose $u = \sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})\ y_i^2$.

D'après (53) et (53.1),

$$K_s(u) = 2^{s-1}\ (s-1)\ ! \sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^s \quad . \tag{55}$$

D'après (40), (41), (52), (54) et (55), on obtient directement:

$$K_1\,(I_x^{CO}) = \mu_1'\,(I_x^{CO}) = \frac{1}{(N-1)} \sum_{i=1}^{N-1} \upsilon_i = \bar{\upsilon} \quad ,$$

$$(56) \qquad K_2(r) = \mu_2(r) = \frac{2 \sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2}{(N-1)\,(N+1)} \quad ,$$

$$K_3(r) = \mu_3(r) = \frac{8 \sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^3}{(N-1)\,(N+1)\,(N+3)} \quad .$$

Par (41), $K_4(r) = \mu_4(r) - 3\,\mu_2^2(r)$, et comme $E(r)^p = E(u)^p/E(v')^p$

$$\mu_4(r) = \frac{\mu_{4(u)}}{\mu_4(v')} = \frac{K_4(u) + 3\,\mu_2^2(u)}{(N^2-1)\,(N+3)\,(N+5)} = \frac{48 \sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^4 + 12 \left[\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2\right]^2}{(N^2-1)\,(N+3)\,(N+5)}$$

d'où :

$$K_4(r) = \frac{48 \sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^4}{(N^2-1)\,(N+3)\,(N+5)} + \frac{12 \left[\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2\right]^2}{(N^2-1)\,(N+3)\,(N+5)} - \frac{12 \left[\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2\right]^2}{(N^2-1)^2} \quad .$$

On obtient finalement[1] :

$$(56.1) \qquad K_4(r) = \frac{48 \sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^4}{(N^2-1)\,(N+3)\,(N+5)} - \frac{96\,(N+2) \left[\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2\right]^2}{(N^2-1)^2\,(N+3)\,(N+5)} \quad .$$

---

1. $12 \left[\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2\right]^2 \cdot \left[\frac{1}{(N^2-1)\,(N+3)\,(N+5)} - \frac{1}{(N^2-1)^2}\right]$

$= 12 \left[\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2\right]^2 \cdot \left[\frac{(N^2-1) - (N+3)\,(N+5)}{(N^2-1)^2\,(N+3)\,(N+5)}\right]$

$= 12 \left[\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2\right]^2 \cdot \left[\frac{N^2-1 - N^2 - 8N - 15}{(N^2-1)^2\,(N+3)\,(N+5)}\right]$

$= 12 \left[\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^2\right]^2 \cdot (-8) \left[\frac{(N+2)}{(N^2-1)^2\,(N+3)\,(N+5)}\right]$.

Grâce au théorème 1 et à l'interrelation entre les moments et les cumulants, par (45), (46) et (47), $r = I_x^{CO} - E(I_x^{CO})$ aura une distribution asymptotique normale si, sous l'hypothèse que $K_2(r)$ converge vers une quantité finie[1] :

$$\frac{K_j(r)}{[K_2(r)]^{j/2}} \tag{57}$$

est $o(1)$[2] pour $j = 3,4,\ldots$.

Implication

Soit $\sum = \sum_{i=1}^{N-1}$.

$$K_3(r) = \frac{2^{6/2} \sum (\upsilon_i - \bar{\upsilon})^3}{(N^2-1)\ (N+3)} \quad ,$$

$$K_2^{3/2}(r) = \frac{2^{3/2} \left[\sum (\upsilon_i - \bar{\upsilon})^2\right]^{3/2}}{(N^2-1)\ (N^2-1)^{1/2}} \quad , \tag{58}$$

d'où

$$\frac{K_3(r)}{K_2(r)^{3/2}} = \frac{2^{6/2} \sum (\upsilon_i - \bar{\upsilon})^3 \cdot (N^2-1)^{1/2}}{2^{3/2} \left[\sum (\upsilon_i - \bar{\upsilon})^2\right]^{3/2} (N+3)} \quad ,$$

$$= \frac{2^{3/2} \cdot (N^2-1)^{1/2}}{(N+3)} \cdot \frac{\sum (\upsilon_i - \bar{\upsilon})^3}{\left[\sum (\upsilon_i - \bar{\upsilon})^2\right]^{3/2}} \quad .$$

---

[1]. Formellement on devrait parler de $K_{N2}(r) \xrightarrow{p} [0 < K_2(r) < \infty]$ et en (57) de $\frac{K_{Nj}(r)}{K_{N2}(r)^{j/2}}$. Pour simplifier la notation, on omet le N.

[2]. h est $o(N)$ si $\lim_{N\to\infty} \left(\frac{h}{N}\right) = 0$

et $O(N)$ si $\lim_{N\to\infty} \frac{h}{N} =$ constante $\neq 0$.

Il suffit alors que[1] :

$$(59) \qquad \frac{\sum (v_i - \bar{v})^3}{[\sum (v_i - \bar{v})^2]^{3/2}} \quad , \quad \text{soit } o(1) \ .$$

De la même façon, il faudra que :

$$(60) \qquad \frac{\sum (v_i - \bar{v})^4}{[\sum (v_i - \bar{v}^2]^2} \quad , \quad \text{soit } o(1) \ ,$$

et de façon générale :

$$(61) \qquad \frac{\sum (v_i - \bar{v})^j}{[\sum (v_i - \bar{v})^2]^{j/2}} \quad , \quad \text{soit } o(1) \quad \text{pour } j \geq 3 \ .$$

(voir Cliff et Ord (1981), p. 48).

---

1. Soit $h(N) = 2^{3/2} \cdot \sqrt{\frac{(N^2 - 1)}{(N + 3)^2}} = 2^{3/2} \cdot \sqrt{\frac{N^2}{(N + 3)^2} - \frac{1}{(N + 3)^2}}$

$$\lim_{N \to \infty} h(N) = \lim_{N \to \infty} 2^{3/2} \sqrt{\frac{\frac{1}{N^2 + 6N + 9}}{N^2} - \frac{1}{(N + 3)^2}} = 2^{3/2} \cdot 1 = 2^{3/2} \ .$$

On peut écrire cette expression en terme des N racines λ_i de W.

Pour que I_x^{CO} soit asymptotiquement normal, il suffit que[1] :

$$(62) \qquad \frac{\sum_{i=1}^{N} \lambda_i^j}{\left[\sum_{i=1}^{N} \lambda_i^2\right]^{j/2}} = \frac{tr\,(W^j)}{tr\,(W^2)^{j/2}} \quad , \qquad \text{soit } o(1) \qquad j \geq 3 \quad .$$

---

[1]. Soit A une matrice quelconque d'ordre (N×N).

(A.1) $tr\,(A) = \sum_{i=1}^{N} \theta_i$ où θ_i est la ième racine caractéristique de A,

$tr\,(A^j) = \sum_{i=1}^{N} \theta_i^j \quad \forall j$ (voir Durbin et Watson (1950), p. 420) .

Comme la matrice symétrique W de cette section est telle que $i'Wi = N$, on obtient :

$$(A.2) \qquad \sum_{i=1}^{N-1} \upsilon_i = tr\,(\bar{T}) = tr\,(W\,M) = tr\,(W) - \frac{1}{N}\, i'Wi = 0 - 1 = -1 \quad .$$

Comme $\sum_{i=1}^{N-1} \upsilon_i = -1$, alors $\lim_{N\to\infty} \bar{\upsilon} = \lim_{N\to\infty} \frac{-1}{N-1} = 0$.

D'où étudier $\sum_{i=1}^{N-1} (\upsilon_i - \bar{\upsilon})^s$ est asymptotiquement équivalent à étudier $\sum_{i=1}^{N-1} \upsilon_i^s$,

car le terme $\bar{\upsilon}$ est, on vient de le voir, négligeable. Comme il est facile de vérifier qu'asymptotiquement $M = [I_N - \frac{1}{N}\, ii'] \to I_N$, alors étudier $\bar{T} = MWM$ est équivalent à étudier W.

Comme

$$(A.3) \qquad \sum_{i=1}^{N-1} \upsilon_i^s = tr\,(\bar{T}^s) \quad ,$$

par la dernière constation, on a :

$$tr\,(\bar{T}^s) \simeq tr\,(W^s) = \sum_{i=1}^{N} \lambda_i^s \quad .$$

C.Q.F.A.

Pour Cliff et Ord (1981), cette condition est peu contraignante et est généralement respectée. Comme condition suffisante mais non nécessaire, ils présentent :

$$\max_i \frac{\lambda_i^2}{2 \text{ tr } (W^2)} \quad \text{doit être} \quad 0\,(N^{-1/2})$$

Remarque

1. Pour la modification de Z (I_x^{CO}) dans de petits échantillons ($n < 50$), consulter Cliff et Ord (1981), section 2.5 ou Cliff et Ord (1971).

1.6.4) *Généralisation du test I_x^{CO} pour un ordre g*

Grâce à (15), (21) et (22), on a :

$$(63) \qquad I_x^{CO}(g) = \frac{N}{S_0(g)} \cdot \frac{z' W(g) z}{z'z} \quad ,$$

car $S_0(g) = i' W(g) i$, avec

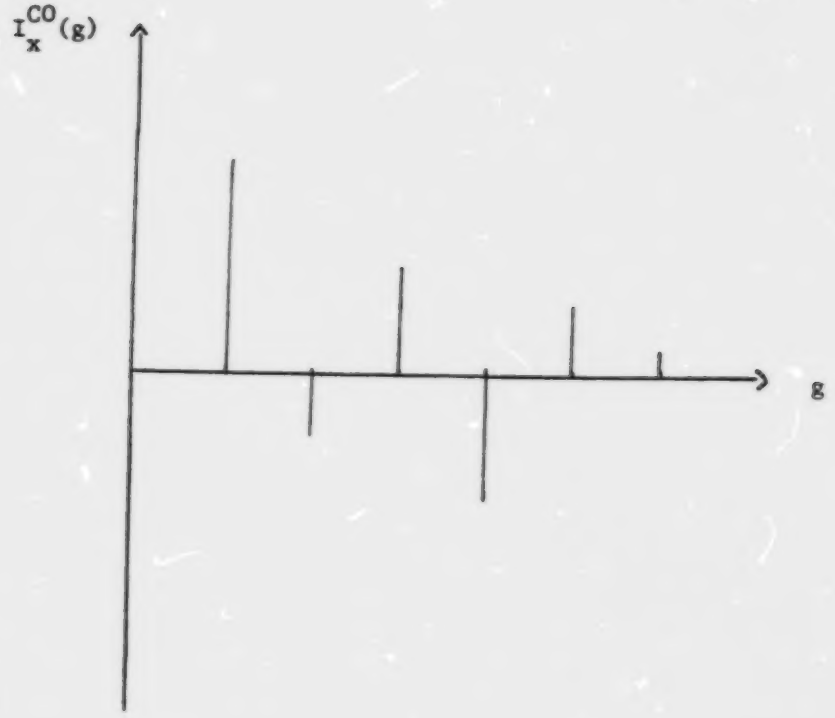
$$(64) \qquad E(I_x^{CO}(g)) = -\frac{1}{(N-1)} \; \forall g \quad ,$$

et

$$(65) \qquad E[(I_x^{CO}(g))^2] = \frac{N^2}{S_0^2(g) \cdot (N^2-1)} \cdot [2 \text{ tr } (W(g) A W(g) A) + \text{tr}^2 (W(g) A)] \; .$$

La formule (63) permet de tracer le corrélogramme spatial.

Exemple



Jusqu'à dernièrement, on testait la signification, dans le corrélogramme, grâce aux moments des distributions marginales évalués par (64) et (65).

Oden (1984) propose une mesure qui permet de tester la signification jointe du corrélogramme. Pour un nombre élevé G de retards, cette statistique peut devenir longue à calculer. Pour plus d'informations, consulter cette référence.

2) L'autocorrélation spatiale dans les résidus

2.1) Les résidus MCO

Soit le modèle suivant :

$$(66) \qquad \underset{(N\times 1)}{y} = \underset{(N\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(N\times 1)}{\varepsilon}\ , \qquad \varepsilon \sim N(0, \sigma^2 I) \quad .$$

Ici, chaque ε_i représente la réalisation dans la zone i du terme d'erreur. On a absence d'autocorrélation spatiale lorsque $F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N) = \prod_{i=1}^{N} F(\varepsilon_i)$. Comme $\varepsilon \sim N$, cette condition est respectée lorsque $E(\varepsilon\varepsilon') = \sigma^2 I$. Comme extension de la section précédente, Cliff et Ord proposent de tester l'autocorrélation spatiale, dans les résidus, avec la mesure suivante : $I_\varepsilon^{CO} = \frac{N}{S_o} \cdot \frac{\varepsilon' W \varepsilon}{\varepsilon'\varepsilon}$. Toutefois, ε_i est en réalité non observable. Alors plusieurs suggèrent de remplacer ε par le vecteur d'erreur e (erreur du MCO). On le rappelle

$$e = y - X\hat{\beta} = y - X(X'X)^{-1} X' y = My = M\varepsilon \quad ,$$

car $MX = 0$, où $M = [I_N - X(X'X)^{-1} X']$.

e est : linéaire car $e = My = M\varepsilon$

: non biaisé car $E(e) = ME(\varepsilon) = 0$,

: à variance minimum car : pour tout vecteur w linéaire en y et non biaisé $E(w - \varepsilon) = 0$, $V(w - \varepsilon)$ excède $V(e-\varepsilon)$ par une matrice semi-définie positive.

(voir théorème 5.1, Theil (1971)).

Soulignons de plus que même si, sous H_0, $E(\varepsilon\varepsilon') = \sigma^2 I$, $E(ee') = E(M\varepsilon\varepsilon'M) = \sigma^2 M$. Donc, sous l'hypothèse de non autocorrélation des erreurs ε_i, les éléments e_i de e sont correlés. Le test d'autocorrélation spatiale sur les résidus observés s'écrit :

$$I_{MCO}^{CO} = \frac{N}{S_0} \frac{e'We}{e'e} \quad . \tag{67}$$

Pour les mêmes raisons qu'à la section 1.6.1.1), le théorème Pitman-Koopmans s'applique. Les moments de distributions de I_{MCO}^{CO} sont évalués à l'appendice II, 3). Comme résultats finaux, on a[1] :

$$E(I_{MCO}^{CO}) = \frac{N}{S_0} \cdot \frac{tr\,(W M)}{(N-K)} \quad , \tag{68}$$

$$E(I_{MCO}^{CO^2}) = \frac{N^2}{S_0^2} \frac{[2\ tr\,(WM\,WM) + tr^2\,(WM)]}{(N-K)\ (N-K+2)} \quad , \tag{69}$$

et

$$V(I_{MCO}^{CO}) = E(I_{MCO}^{CO^2}) - E^2(I_{MCO}^{CO}) \quad , \tag{70}$$

où $M = [I - X(X'X)^{-1} X']$.

---

[1]. On peut vérifier que ces formules sont équivalentes à celles présentées au chapitre 8 de Cliff et Ord (1981).

Pour I_x^{CO}, on avait comme expression équivalente à (68), $-(N-1)^{-1}$, alors que maintenant l'espérance mathématique dépend d'une façon complexe de X et donc de l'échantillon, ce qui rend la mesure (67) moins attrayante.

Afin d'éviter l'inconvénient causé par les covariances non nulles de e sous H_0, des procédures LUS[1], visant à donner aux résidus estimés une matrice variance-covariance scalaire, ont été suggérées. Avant de les étudier, parlons de la normalité asymptotique de I_{MCO}^{CO}. De cette façon, le test d'autocorrélation spatiale consistera à comparer la valeur de la statistique standardisée à une N(0,1).

2.1.1) *Normalité asymptotique*

(67) peut s'écrire :

$$(71) \qquad I_{MCO}^{CO} = \frac{\varepsilon' M' W M \varepsilon}{\varepsilon' M \varepsilon} \quad .$$

Comme I_{MCO}^{CO} est une mesure invariante par rapport à l'échelle, alors, sans perte de généralité, on peut poser $\sigma^2 = 1$, d'où $\varepsilon \sim N(0,I)$. L'expression (71) est donc similaire à la formule (25) de la section 1.6.3), sauf que maintenant $M = [I - X(X'X)^{-1} X']$ au lieu de $[I - \frac{1}{N} ii']$. Malgré cela,

---

[1]. LUS : linear unbiased with scalar covariance matrix.

pour Cliff et Ord (1981), les mêmes conclusions qu'en 1.6.3) s'appliquent. Pour avoir la normalité, d'après l'équation (62), il faut que :

$$(72) \qquad \frac{\text{tr } W^j}{[\text{tr } W^2]^{j/2}}, \qquad \text{soit } o(1) \qquad j \geq 3,$$

laquelle condition est, d'après Cliff et Ord, généralement respectée.

2.2) Procédures alternatives

Lorsqu'on utilise e comme estimateur de ε, la présence d'une matrice variance-covariance non scalaire nous a amené des complications lors de l'évaluation des moments. Pour contrevenir à cet inconvénient, on peut se tourner vers la classe des estimateurs $\hat{e}$ de ε qui possèdent une matrice variance-covariance scalaire (σ^2 I). La classe LUS, regroupe ces divers estimateurs.

La classe LUS

Pour être LUS (linear unbiased with scalar covariance matrix), $\hat{e}$ doit respecter :

- linéarité

$$(73) \qquad \hat{e} = Cy$$

- non biais

(74) $\hat{e} = C(X\beta + \varepsilon) = C\varepsilon$,

d'où $E(\hat{e}) = 0$, lorsque

(75) $CX = 0$.

Alors (75) impose K combinaisons linéaires sur ε. Theil (1971), p. 203, montre que si X est de rang complet, C est d'ordre $(N-K) \times N$. Donc, par (73), on ne pourra calculer que (N-K) éléments $\hat{e}_i$ de $\hat{e}$.

- matrice v.c. scalaire

Par (74), on a :

(76) $V(\hat{e}) = \sigma^2 CC' = \sigma^2 I_{(N-K)}$,

(77) si $I_{(N-K)} = CC'$.

De façon générale, pour estimer $\hat{e}$, il faut toujours délaisser K observations de l'échantillon. Ces dernières forment alors la base (Theil, p. 203).

Parmi les membres de la classe LUS, il existe des estimateurs e^B qui minimisent $E(\hat{e}_1 - \varepsilon_1)' (\hat{e}_1 - \varepsilon_1)$ (voir Theil, p. 212)[1]. e^B est dit résidu BLUS.

---

[1]. Pour la définiton de ε_1 et $\hat{e}_1$, voir la page suivante.

Theil (1971) propose de trouver e^B comme suit :

résidus BLUS

Partageons les observations de telle façon que les K membres de la base soient au début. Alors (66) devient :

$$(78) \quad \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} X_0 \\ X_1 \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \end{bmatrix} = \begin{bmatrix} X_0 \\ X_1 \end{bmatrix} \hat{\beta}_{MCO} + \begin{bmatrix} e_0 \\ e_1 \end{bmatrix} ,$$

où y_0, X_0 et e_0 appartiennent à la base. On peut toujours trouver un X_0 $_{(K\times K)}$ tel que X_0^{-1} existe. Alors Theil (1971, théorème 5.4) établit que :

$$(79) \quad \underset{([N-K]\times 1)}{e^B} = \underset{([N-K]\times 1)}{e_1} - X_1 X_0^{-1} \left[\sum_{j=1}^{J} \frac{d_j}{1+d_j} q_j q_j' \right] e_0 ,$$

d'où e^B se trouve à partir des résidus MCO e, lesquels ont été partagés. d_j est la valeur absolue de d_j^2, où d_j^2 est la racine caractéristique j de $X_0 (X'X)^{-1} X_0'$, et q_j le vecteur caractéristique correspondant, où on a J racines $d_j < 1$.

Comme aspect positif, les moments de distributions du test d'autocorrélation sont simplifiés[1], par contre (79) exige le choix arbitraire de la base, et le calcul de racines caractéristiques, etc.... . Alors, pour faciliter la procédure de calcul des résidus estimateurs de ε, Phillips et Harvey (1974), en abandonnant la propriété de meilleur estimateur (Best), proposent une estimation récursive simple pour estimer ê (membre de la classe LUS). Appelons cet estimateur e^R (RELUS). Pour les aspects techniques, consulter Brandsma et Ketellaper (1979b) et Bartels et Hordijk (1977).

---

[1]. Voir équations (80) à (82) suivantes.

Avec des données spatiales, le choix de la base demeure un problème. Ces derniers auteurs proposent de délaisser les K zones les moins contiguës. Comme e^B et e^R respectent tous deux (73) à (77), le test d'autocorrélation spatiale sur les résidus devient donc[1] :

$$(80) \qquad I_{\hat{e}}^{CO} = \frac{(N-K)}{S_0^*} \frac{\hat{e}' W_0 \hat{e}}{\hat{e}'\hat{e}} \quad ,$$

où W_0 est la matrice W dont les colonnes et lignes des membres de la base ont été enlevées, et où

$$S_0^* = i' W_0 i \quad ; \qquad \text{où ici } i_{(N-K)\times 1} \quad .$$

La normalité asymptotique découle de I_{MCO}^{CO} et les moments s'expriment comme suit :

$$(81) \qquad E(I_{\hat{e}}^{CO}) = 0 \quad ,$$

$$(82) \qquad E(I_{\hat{e}}^{CO})^2 = \frac{(N-K)}{(N-K+2)} \cdot \frac{2 \text{ tr } (W_0^2)}{S_0^{*2}} \quad .$$

Donc, comparativement à (68) et (69), les résidus LUS redonnent la propriété d'espérance mathématique constante à la mesure d'autocorrélation spatiale.

On a maintenant potentiellement, trois méthodes pour tester l'autocorrélation spatiale dans les résidus. La première consiste à estimer ε par e (résidu du MCO) et appliquer les équations (67) à (69) pour, par la suite, comparer la statistique standardisée à une N(0,1). Avec les deux autres,

---

[1]. Pour la dérivation des moments, consulter l'appendice II, 3).

appliquer selon le cas e^B ou e^R dans les formules (80) à (82) et comparer la statistique standardisée à une N(0,1). Quelle approche est-il préférable de prendre ? De façon raisonnable, on suggère de choisir le test qui s'est le mieux comporté lors des expériences empiriques. Autrement dit, retenir le test qui montre la plus forte puissance empirique[1].

Pour ce faire, présentons maintenant les résultats des études de Monté-Carlo par Brandsma et Ketellaper (1979b) et Bartels et Hordijk (1977).

2.3) Expériences de Monté-Carlo

Avec la notation établie au chapitre I, un modèle autorégressif d'ordre 1 sur les résidus peut s'écrire[2] :

(83) $\quad y = X\beta + \varepsilon \quad ,$

(84) $\quad \varepsilon = \rho W \varepsilon + u \quad , \qquad u \sim N(0, \sigma^2 I) \quad ,$

où W est la matrice de contiguïté.

Si on se limite à l'autocorrélation positive ($\rho > 0$), le problème se formule aussi comme :

sous H_0 : $\quad E(\varepsilon\varepsilon') = \sigma^2 I \quad ,$

sous H_1 : $\quad \varepsilon = \rho W \varepsilon + u \quad , \qquad$ où $\rho > 0.$

---

[1]. La puissance d'un test s'écrit 1 - P (type II). C'est donc la probabilité de rejeter H_0 lorsque H_0 est faux. Empiriquement, c'est de donner le pourcentage de fois que H_0 est rejeté, alors qu'on a généré des observations sous l'hypothèse H_1.

[2]. Voir aussi le chapitre III.

Brandsma et Ketellaper, en plus de comparer la puissance des tests d'autocorrélation utilisant e, e^B et e^R, veulent analyser le comportement du test du ratio des vraisemblances (Likelihood Ration Test).

2.3.1) *Le ratio des vraisemblances*

Sous H_0, la fonction de vraisemblance de (83) s'écrit :

$$(85) \qquad \mathcal{L}(\beta, \sigma^2, 0) = (2\pi)^{-N/2}\, \sigma^{-N} \exp\{-\frac{1}{2\sigma^2}(y - X\beta)'(Y - X\beta)\} \ ,$$

et sous H_1, on obtient :

$$(86) \qquad \mathcal{L}(\beta, \sigma^2, \rho) = (2\pi)^{-N/2}\, \sigma^{-N}\, |P| \exp\{-\frac{1}{2\sigma^2}(y - X\beta)'\, V^{-1}\, (y - X\beta)\} \ ,$$

où $P = (I - \rho W)$ et $V^{-1} = P'P$.

Sous l'hypothèse nulle $-2 \ln \lambda \underset{N\to\infty}{\sim} \chi^2_1$; où λ se définit comme le ratio des vraisemblances :

$$(87) \qquad \lambda = \frac{\max \mathcal{L}(\beta, \sigma^2, 0)}{\max \mathcal{L}(\beta, \sigma^2, \rho)} \ , \qquad \lambda < 1 \ .$$

Afin d'évaluer la puissance de ces divers tests, ces deux groupes d'auteurs étudient les mêmes échantillons et génèrent des erreurs autocorrelées ε selon (84)[1].

Suite aux deux études[2], il semble que la procédure la plus intéressante consiste à tester l'autocorrélation spatiale à partir des résidus MCO, en utilisant les formules (68) et (69). De façon générale, Brandsma et Ketellaper classent comme second-best la procédure avec résidus BLUS.

---

[1]. Pour des détails sur les simulations, consulter ces références.

[2]. Le nombre de régions de l'échantillon est 39, et le nombre de variables explicatives passe de 3 à 6.

CHAPITRE III

ESTIMATION DU SAR(1) RÉSIDUEL

Au cours de ce chapitre, on étudiera les méthodes déjà suggérées qui concernent l'estimation du modèle SAR(1) suivant :

(1) $y = X\beta + \varepsilon$,

(2) $\varepsilon = \rho W \varepsilon + u$, $u \sim N(0, \sigma^2 I)$,

où, on le rappelle, W est la matrice (connue) de contiguïté d'ordre 1, y, ε et u sont des vecteurs aléatoires d'ordre (N×1) et X une matrice fixe (N×K).

Note : Dans le traitement théorique du modèle SAR(1) résiduel, on fait l'hypothèse d'absence d'hétéroscédasticité résiduelle. Pour le cas hétéro-spatial, c'est-à-dire, le cas où il se produit simultanément des problèmes d'hétéroscédasticité et d'autocorrélation, les résultats établis aux sections 1) et 4) du chapitre X, permettent de proposer la procédure en deux étapes convergente suivante. Dans un premier temps, modéliser l'hétéroscédasticité selon la procédure convergence développée à la section 4.1) du chapitre X. En second lieu, appliquer sur le modèle homoscédastique transformé, la méthodologie SAR(1) du présent chapitre.

De la même façon qu'avec des données de séries temporelles, afin d'estimer ρ et β, on peut suggérer l'utilisation de la méthode en deux étapes de Durbin (1960), l'application de la procédure itérative de Cochrane-Orcutt ou bien l'estimation conjointe du vecteur $[\beta, \rho]'$ par la méthode du maximum de vraisem-

blance. Selon les deux premières méthodes, on procède à l'application des moindres carrés ordinaires sur des réécritures des équations (1) et (2)[1]. Pour des données temporelles, ces méthodes produisent des estimateurs de β et ρ qui sont convergents et asymptotiquement efficaces. Ils ont en effet les mêmes propriétés asymptotiques que les estimateurs du maximum de vraisemblance. Pour un rappel des concepts asymptotiques d'estimateurs, consulter l'appendice III.

Pour des données spatiales, on montrera que généralement, à l'équation (2), l'estimation de ρ par la méthode des moindres carrés ordinaires produit un estimateur $\hat{\rho}$ non convergent.

Avant d'entrer dans les détails d'estimation, décrivons les schémas de dépendance et les caractéristiques de certaines configurations spatiales qui serviront à étayer les démonstrations de ce chapitre. Afin de respecter l'ordre dans lequel ils ont été introduits au chapitre I, nous commencons par le modèle bilatéral unidimensionnel "line transect". Par la suite, nous considérons le modèle autorégressif quadrilataire d'ordre 1 et le modèle spatial en vague "wave model". Comme on le sait, ces trois modèles sont applicables à des configurations spatiales rectangulaires. Pour terminer, nous introduisons trois schémas de dépendance lesquels représentent des cas extrêmes dans la modélisation de configurations spatiales irrégulières.

---

[1]. Voir Johnston (1984), chapitre 8.

1) Quelques schémas de dépendance spatiale dans les résidus

1.1) Le modèle spatial bilatéral unidimensionnel: cas ①

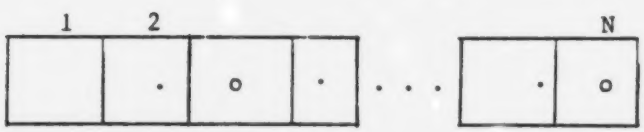
D'après ce modèle, les équations (1) et (2) s'écrivent :

$$(3) \qquad y_i = \sum_{k=1}^{K} x_i^k \beta_k + \varepsilon_i \quad , \qquad i = 1,\ldots,N$$

$$(4) \qquad \varepsilon_i = \frac{\rho}{2} (\varepsilon_{i-1} + \varepsilon_{i+1}) + u_i \quad . \qquad u_i \sim BB\ (0, \sigma^2)$$

Nous divisons ρ par 2 pour que le modèle soit stationnaire pour $|\rho| < 1$. Le schéma de dépendance se retrouve à la figure 1.

FIGURE 1



soit C la matrice de contiguïté suivante :

$$(5) \qquad C = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots\ldots & 0 \\ 1 & 0 & 1 & 0 & \ldots\ldots & 0 \\ 0 & 1 & 0 & 1 & \ldots\ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \ldots\ldots & 0 \end{bmatrix} ,$$

c'est-à-dire, pour $i \neq 1$ ou N,

$$(6) \qquad c_{ij} = \begin{cases} 1 & \text{si } i-1 = j \text{ ou } i+1 = j \\ 0 & \text{ailleurs .} \end{cases}$$

Matriciellement, pour $W = 1/2\ C$, les équations (3) et (4) s'écrivent :

$$y = X\beta + \varepsilon \quad , \tag{7}$$

$$\varepsilon = \rho W \varepsilon + u \quad . \qquad u \sim BB\ (0, \sigma^2 I_N) \ . \tag{8}$$

Soit W formé grâce à l'équation (5), Ord (1975) obtient que $d_N = \det [\lambda I_N - W]$ définit un polynôme d'ordre N qui respecte la récurrence suivante :

$$d_N = 2\lambda\, d_{N-1} - d_{N-2} \quad . \tag{9}$$

Comme d'après le théorème 1 du présent chapitre[1] $-1 \leq \lambda \leq 1$, l'équation (9) représente la relation de récurrence du polynôme de Chebyshev du second type. Cette information permet de déterminer que les zéros de $d_N = \det [\lambda I_N - W]$ sont[2] :

$$\lambda_i = \cos\left(\frac{\pi i}{N+1}\right) \quad . \qquad i = 1,2,\ldots,N. \tag{10}$$

1.2) Le modèle autorégressif quadrilataire d'ordre 1 : QA(1): cas ②

D'après les résultats du chapitre 1, ce modèle stationnaire pour $|\rho| < 1$ devrait s'écrire :

$$y_{r,s} = \sum_{k=1}^{K} x^k_{r,s}\, \beta_k + \xi_{r,s} \quad , \qquad \begin{matrix} r = 1,\ldots,P \\ s = 1,\ldots,Q \end{matrix} \tag{11}$$

$$\xi_{r,s} = \frac{\rho}{4} (\xi_{r-1,s} + \xi_{r+1,s} + \xi_{r,s-1} + \xi_{r,s+1}) + \zeta_{r,s} \ , \quad \zeta_{r,s} \sim BB\ (0, \sigma^2) \tag{12}$$

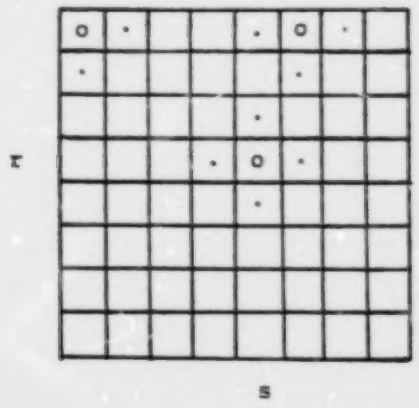
---

1. $|\lambda| \leq \max_i \sum |w_{ij}| = 1$

2. Voir Abramowitz et Stegun, 1972, section 22.16.

A la figure 2, on présente le schéma des dépendances entre les résidus $\xi_{r,s}$ de ce modèle.

FIGURE 2



Soit C_1 la matrice C de l'équation (5) de dimension (P×P) et C_2 la même matrice C de dimension (Q×Q), l'équation (12) matriciellement s'écrit :

$$\xi = \frac{\rho}{4} (C_1 \xi + \xi C_2) + \zeta \quad , \tag{13}$$

où ξ et ζ sont des matrices de dimension (P×Q).

Lorsqu'on renumérote les zones de la région selon les lignes comme suit :

1	2	3	Q
Q+1	Q+2		2Q
2Q+1	.....		3Q
⋮			⋮
(P-1)Q+1	.....		PQ

ceci revient à appliquer à l'équation (13) la transformation vec[1]. On obtient :

$$(14) \qquad \underset{(PQ\times 1)}{\text{vec}\ (\xi)} = \frac{\rho}{4}\left[(C_1 \otimes I_Q) + (I_P \otimes C_2)\right] \text{vec}\ (\xi) + \text{vec}\ (\zeta) \quad .$$

Soit $N = PQ$, $\underset{(N\times 1)}{\varepsilon} = \text{vec}\ (\xi)$, $u = \text{vec}\ (\zeta)$ et

$$\underset{(N\times N)}{W} = \frac{1}{4}\left[(C_1 \otimes I_Q) + (I_P \otimes C_2)\right]$$

$$(15) \qquad = \frac{1}{4}\begin{bmatrix} C_2 & I_Q & 0_Q & 0_Q & \cdots\cdots & 0_Q \\ I_Q & C_2 & I_Q & 0_Q & \cdots\cdots & 0_Q \\ 0_Q & I_Q & C_2 & I_Q & \cdots\cdots & 0_Q \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0_Q & 0_Q & 0_Q & 0_Q & \cdots\cdots & C_2 \end{bmatrix}$$

en effectuant le même type de transformation sur l'équation (11), le QA(1) s'écrit :

$$(16) \qquad \begin{aligned} y_i &= \sum_k^K x_i^k \beta_k + \varepsilon_i \quad , \\ \varepsilon_i &= \rho \sum_{j=1}^N w_{ij}\, \varepsilon_j + u_i \quad , \end{aligned} \qquad i = 1,2,\ldots,N$$

ou matriciellement :

$$(17) \qquad \begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &= \rho W \varepsilon + u \end{aligned} \qquad u \sim BB\ (0, \sigma^2 I_N) \quad .$$

---

[1]. $\text{vec}\ (ABC) = (A \otimes C')\ \text{vec}\ (B)$

L'équation (15) permet à Ord (1975) d'établir que les racines $\lambda_{r,s}$ de W sont :

$$\lambda_{r,s} = \frac{1}{2}\left[\cos\left(\frac{r\pi}{P+1}\right) + \cos\left(\frac{s\pi}{Q+1}\right)\right] \quad , \qquad \begin{matrix} r = 1,\ldots,P \\ s = 1,\ldots,Q \end{matrix} \quad . \tag{18}$$

Pour avoir λ_i, $i = 1,\ldots,N$, il suffit d'appliquer l'opérateur vec à la matrice $\underset{(P\times Q)}{\Lambda}$ des racines $\lambda_{r,s}$. On a donc :

$$\underset{(N\times 1)}{\lambda} = \text{vec}\,(\Lambda) \quad . \tag{19}$$

1.3) Le modèle autorégressif spatial en vague: cas ③

Le schéma autorégressif des résidus de ce modèle s'écrit :

$$\xi_{r,s} = \frac{\rho}{2}\left(\xi_{r-1,s} + \xi_{r,s-1}\right) + \zeta_{r,s} \quad , \tag{20}$$

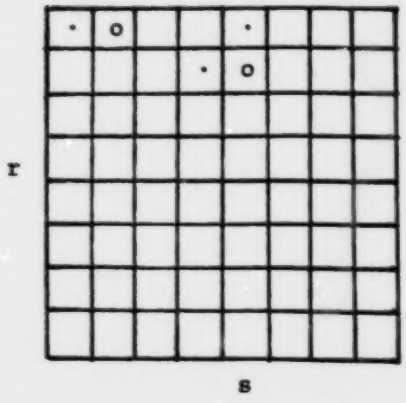
avec stationnarité lorsque $|\rho| < 1$. Grâce à

$$\underset{(P\times P)}{C_1} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots\ 1 & 0 \end{bmatrix} \quad \text{et} \quad \underset{(Q\times Q)}{C_2} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

l'expression (20) dont le schéma de dépendance est à la figure 3 s'écrit matriciellement :

$$\underset{(P\times Q)}{\xi} = \frac{\rho}{2}\left(C_1\,\xi + \xi\,C_2\right) + \underset{(P\times Q)}{\zeta} \quad , \tag{21}$$

FIGURE 3



Si on vectorise (21) de la même façon qu'en 1.2), on obtient :

$$\varepsilon = \frac{\rho}{2} \left[(C_1 \otimes I_Q) + (I_P \otimes C_2') \right] \varepsilon + u \quad . \tag{22}$$

Le modèle autorégressif en vague s'écrit donc matriciellement sous la forme :

$$\begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &= \rho W \varepsilon + u \quad , \end{aligned} \tag{23}$$

avec

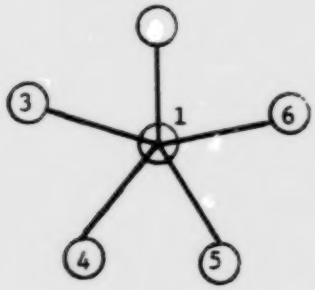
$$(24) \qquad \underset{(N\times N)}{W} = \begin{bmatrix} C_2' & 0_Q & 0_Q & \cdots & 0_Q \\ I_Q & C_2' & 0_Q & \cdots & 0_Q \\ 0_Q & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0_Q \\ 0_Q & \cdots & I_Q & & C_2' \end{bmatrix} .$$

Ainsi, pour ce modèle, W est une matrice triangulaire par le bas dont la diagonale principale ne contient que des termes nuls. Les racines caractéristiques λ_i, i = 1,...,N de W sont donc ici toutes nulles.

1.4) Configuration spatiale en étoile: cas ④

A la région de la figure 4, nous remarquons que la zone 1 est dominante en ce sens que toutes les zones sont contiguës à cette zone sans l'être entre elles.

FIGURE 4



La matrice de contiguïté pour un tel phénomène réalisé sur N zones s'écrit :

$$(25) \qquad \underset{(N\times N)}{W} = \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} .$$

Evidemment, comme cette formulation concerne une configuration spatiale irrégulière (non rectangulaire), nous pouvons directement exprimer ce modèle sous la forme des équations (1) et (2).

Comme racines caractéristiques de W, on obtient :

$$\lambda_i = 0 \qquad i = 1,2,\ldots,N-2,$$

$$(26) \qquad \lambda_{N-1} = -\sqrt{(N-1)}$$

$$\lambda_N = \sqrt{(N-1)} \quad .$$

Preuve

$$0 = \begin{vmatrix} -\lambda & 1 & 1 \ldots & 1 \\ 1 & -\lambda & 0 \ldots & 0 \\ 1 & 0 & -\lambda \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 \ldots & -\lambda \end{vmatrix} = \begin{vmatrix} -\lambda & 1 & 1 \ldots & 1 \\ 1 & -\lambda & 0 \ldots & 0 \\ 0 & \lambda & -\lambda \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \lambda & 0 & -\lambda \end{vmatrix} = \left|\begin{array}{cc|ccc} -\lambda & (N-1) & 1 & 1 \ldots & 1 \\ 1 & -\lambda & 0 & 0 \ldots & 0 \\ \hline 0 & 0 & -\lambda & 0 \ldots & 0 \\ \vdots & \vdots & 0 & -\lambda \ldots & 0 \\ 0 & 0 & 0 & 0 & -\lambda \end{array}\right|$$

$$0 = \lambda^{(N-2)} \cdot [\lambda^2 - (N-1)] = \lambda^{(N-2)} \; (\lambda - \sqrt{(N-1)}) \; (\lambda + \sqrt{(N-1)})$$

C.Q.F.D.

1.5) Contiguïté spatiale totale : cas ⑤

Ici on considère le cas extrême (non réaliste) que toutes les régions sont contiguës avec $w_{ij} = k \quad \forall i \neq j$. On a donc :

$$W = \begin{bmatrix} 0 & k & k & \ldots\ldots & k \\ k & 0 & k & \ldots\ldots & k \\ k & k & 0 & \ldots\ldots & k \\ \vdots & \vdots & \vdots & & \vdots \\ k & k & k & \ldots\ldots & 0 \end{bmatrix} \qquad (27)$$

Les racines caractéristiques de W sont :

$$\lambda_1 = k\,(N-1)$$
$$\lambda_i = -k \qquad i = 2,\ldots,N \qquad (28)$$

Preuve

$$0 = \begin{vmatrix} -\lambda & k & k & \ldots & k \\ k & -\lambda & k & \ldots & k \\ \vdots & \vdots & \vdots & & \vdots \\ k & k & k & \ldots & -\lambda \end{vmatrix} .$$

En enlevant la première ligne de toutes les autres et en additionnant à la première colonne toutes les autres, on obtient :

$$0 = \begin{vmatrix} -\lambda & k & k & \dots & k \\ k+\lambda & -(k+\lambda) & 0 & \dots & 0 \\ k+\lambda & 0 & -(k+\lambda) & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ k+\lambda & 0 & 0 & & -(k+\lambda) \end{vmatrix} = \begin{vmatrix} -\lambda+(N-1)k & k & k & \dots & k \\ 0 & -(k+\lambda) & 0 & \dots & 0 \\ 0 & 0 & -(k+\lambda) & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & -(k+\lambda) \end{vmatrix},$$

d'où

$$0 = (k+\lambda)^{(N-1)} \, [(N-1)\, k-\lambda] \quad . \tag{29}$$

C.Q.F.D.

1.6) Contiguïté spatiale standard: cas ⑥

Comme sixième configuration spatiale, on inclut ici le modèle spatial avec configuration irrégulière dont le schéma de dépendance est standard, c'est-à-dire le cas où aucune région ne semble dominer dans la structure de contiguïté et où chaque zone est contiguë avec un nombre limité de zones avoisinantes. Plus deux zones sont éloignées et moins la contiguïté entre ces deux zones doit être forte. Comme exemple, se reporter à la figure 7 du chapitre I. Les racines caractéristiques de la matrice W, laquelle dépend de la région étudiée, sont calculées par ordinateur.

On aimerait maintenant traiter des méthodes qui visent l'estimation des paramètres $\underset{(K\times 1)}{\beta}$ et ρ aux équations (1) et (2).

2) Estimation de ρ par la méthode des moindres carrés ordinaires (MCO)

Comme déjà mentionné, les méthodes en 2 étapes de Durbin ou de Cochrane-Orcutt proposent pour l'estimation de β et ρ des estimations MCO sur des réécritures des équations (1) et (2).

Soit l'équation (2) :

$$\varepsilon = \rho W \varepsilon + u \quad , \qquad E(u) = 0 \qquad E(uu') = \sigma^2 I \quad . \tag{30}$$

La minimisation de $u'u = \varepsilon' (I - \rho W)' (I - \rho W)\varepsilon$ par rapport à ρ produit l'estimateur $\hat{\rho}$ suivant :

$$\hat{\rho} = \frac{\varepsilon' W \varepsilon}{\varepsilon' W'W \varepsilon} = \rho + \frac{\varepsilon' W u}{\varepsilon' W'W \varepsilon} \quad . \tag{31}$$

Comme $\hat{\rho}$ est fonction d'un rapport de formes quadratiques produites à partir du vecteur aléatoire ε, les propriétés d'échantillon fini de $\hat{\rho}$ sont difficiles à évaluer. Nous étudierons donc les propriétés asymptotiques de $\hat{\rho}$ [1,2].

Comme

$$\operatorname*{plim}_{N\to\infty} \hat{\rho} = \rho + \frac{\operatorname*{plim}_{N\to\infty} \frac{\varepsilon' W u}{N}}{\operatorname*{plim}_{N\to\infty} \frac{\varepsilon' W'W \varepsilon}{N}} \quad , \tag{32}$$

---

1. Ainsi qu'observé par W.D. Fisher (1971), le concept d'asymptoticité est plus justifiable en temporel qu'en spatial. En effet, il est difficile de se convaincre qu'une même région croîtra en taille, et que de plus, l'ajout des zones en bordure de la région se fera de façon cohérente avec le schéma de dépendance déjà présent entre les zones de la région. L'état de l'art étant ce qu'il est, les simplifications théoriques emmenées par l'utilisation des concepts de grands échantillons nous poussent à nous reléguer malgré la dernière remarque, à l'étude des propriétés asymptotiques des estimateurs.

2. Pour un rappel des propriétés d'estimateurs, consulter l'appendice III.

en posant comme hypothèse :

$$(33) \qquad \operatorname*{plim}_{N\to\infty} \frac{\varepsilon' W' W \varepsilon}{N} = K < \infty \quad ,$$

si

$$(34) \qquad \operatorname*{plim}_{N\to\infty} \frac{\varepsilon' W u}{N} = 0 \quad ,$$

$\hat{\rho}$ est un estimateur convergent de ρ.

Il est difficile d'évaluer de façon générale l'expression (34) car elle dépend de la matrice W laquelle est différente d'une étude à l'autre. Signalons que cette expression est la contrepartie spatiale de la condition qui dit qu'en temporel, pour $y_t = \rho\, y_{t-1} + u_t$, l'estimateur de ρ par la méthode des MCO est justifiable lorsque $\operatorname{plim} \frac{y_{t-1}\, u_t}{N} = 0$. Terme à terme, l'expression (34) devient :

$$(35) \qquad \operatorname*{plim}_{N\to\infty} \frac{\sum_{i=1}^{N} u_i \left(\sum_{j=1}^{N} w_{ij}\, \varepsilon_j \right)}{N} = \operatorname*{plim}_{N\to\infty} \frac{\sum_{i}^{N} \sum_{j}^{N} w_{ij}\, u_i\, \varepsilon_j}{N} \quad .$$

$\hat{\rho}$ est donc convergent lorsque le terme $\sum_{j=1}^{N} w_{ij}\, \varepsilon_j$ est asymptotiquement indépendant de u_i [1]. Comme condition suffisante pour que (34) soit respectée, il faut que :

$$(36) \qquad \lim_{N\to\infty} E \left(\frac{\varepsilon' W u}{N} \right) = 0 \quad ,$$

et

$$(37) \qquad \lim_{N\to\infty} V \left(\frac{\varepsilon' W u}{N} \right) = 0 \quad .$$

---

[1]. Voir Johnston (1984), p. 361.

Lorsque la somme des éléments sur la diagonale de la matrice des covariances croisées $E(W\varepsilon \cdot u)$ qui, comme $\varepsilon = (I - \rho W)^{-1} u$ s'écrit :

$$(38) \qquad E(W (I - \rho W)^{-1} uu') = \sigma^2 W (I - \rho W)^{-1} ,$$

est égal à zéro, (36) est égal à zéro. En effet, on a :

$$\frac{\sigma^2}{N} W(I - \rho W)^{-1} = \frac{1}{N} \left\{ E \begin{bmatrix} \sum_j w_{1j} \varepsilon_j \\ \vdots \\ \sum_j w_{Nj} \varepsilon_j \end{bmatrix} [u_1 \quad u_2 \ \cdots \ u_N] \right\} .$$

Remarque

Lorsque W est triangulaire "cas ③" (disons par le bas), $(I - \rho W)^{-1}$ est aussi triangulaire (par le bas). Par exemple, pour N = 3, on a pour (38) :

$$\begin{bmatrix} 0 & 0 & 0 \\ w_{21} & 0 & 0 \\ w_{31} & w_{32} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ b_{21} & 1 & 0 \\ b_{31} & b_{32} & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ w_{21} & 0 & 0 \\ w_{31}+b_{21}w_{32} & w_{32} & 0 \end{bmatrix} .$$

Ainsi, pour cet exemple spatial avec dépendance unilatérale (W triangulaire), nous obtenons que :

$$(39) \qquad E\left[\left(\sum_j w_{ij} \varepsilon_j\right) \cdot u_h\right] = 0 , \qquad \forall h \geq i .$$

Tout comme pour le modèle AR(1) temporel $\varepsilon_i = \rho \varepsilon_{i-1} + u_i$ avec $E(\varepsilon_{i-1} \cdot u_h) = 0 \ \forall h \geq i$, l'erreur u_h n'est pas correlé avec les observations présentes et passées de la variable explicative endogène retardée.

Comme premier résultat, on a donc que :

$$E\left(\frac{\varepsilon' W u}{N}\right) = \sigma^2 \operatorname{tr}\left(\frac{W (I - \rho W)^{-1}}{N}\right)$$

est égal à zéro lorsque W est triangulaire.

Si on rassemble les résultats précédents, pour un modèle spatial avec dépendance multidirectionnelle[1], il est plus plausible de croire que :

$$\operatorname*{plim}_{N\to\infty} \frac{\varepsilon' W u}{N} = g < \infty$$

où g est un nombre fini différent de zéro. Par exemple, pour le cas ②, il n'est pas raisonnable de croire que $u_{r,s}$ est indépendant de $\varepsilon_{r-1,s} + \varepsilon_{r+1,s} + \varepsilon_{r,s-1} + \varepsilon_{r,s+1}$. En effet, par exemple, $\varepsilon_{r-1,s} = \frac{\rho}{4}(\varepsilon_{r-2,s} + \varepsilon_{r,s} + \varepsilon_{r-1,s-1} + \varepsilon_{r-1,s+1}) + u_{r-1,s}$. Ainsi, $E(u_{r,s} \cdot \varepsilon_{r-1,s}) = 0$ ne fait pas de sens car $\varepsilon_{r-1,s} = f(\varepsilon_{r,s})$, et on sait que $E(\varepsilon_{r,s} \cdot u_{r,s}) \neq 0$. Cette condition suffit à démontrer que dans le cas de dépendance multidirectionnelle, $\operatorname*{plim}_{N\to\infty} \frac{\varepsilon' W u}{N} \neq 0$.

Ainsi, avec (33) respecté, l'estimation de ρ à l'équation (2) par les moindres carrés ordinaires sera généralement non convergente. Pour le cas ③ avec dépendance unilatérale "wave model" comme dans ce cas[2] $\operatorname*{plim}_{N\to\infty} \frac{\varepsilon' W u}{N} = 0$, l'estimation $\hat{\rho}$ des moindres carrés ordinaires est convergente.

---

[1]. Voir les cas ① ② ④ ⑤ et ⑥ de la section 1.

[2]. Ici, on considère que $\lim_{N\to\infty} V\left(\frac{\varepsilon' W u}{N}\right) = \lim_{N\to\infty} \frac{1}{N^2} V(\varepsilon' W u) = 0$.

3) Méthode itérative Cochrane-Orcutt modifiée (Ord (1975))

Dans les deux sections précédentes, on a établi que généralement, pour le modèle

$$y = X\beta + \varepsilon \quad ,$$

$$\varepsilon = \rho W \varepsilon + u \quad , \qquad u \sim N(0, \sigma^2 I_N) \quad ,$$

l'estimation de ρ par la méthode des moindres carrés produit un estimateur $\hat{\rho}$ non convergent. Afin d'obtenir une méthode d'estimation itérative de type Cochrane-Orcutt qui produit des estimateurs β et ρ convergents, Ord (1975) propose d'estimer ρ par une méthode du maximum de vraisemblance[1]. Comme ε est inobservable, il remplace ε par $\tilde{e}$ le vecteur d'erreur MCO. Ceci est justifiable car il est possible de montrer que pour X fixe, chaque $\tilde{e}_t$ converge en distribution vers ε_t, $\forall t$. Comme algorithme, il propose :

1. d'estimer l'équation (1) par les MCO et de calculer le vecteur $\tilde{e} = (y - X\tilde{\beta})$;
2. d'estimer par la méthode du maximum de vraisemblance $\tilde{\rho}$ de $\tilde{e} = \rho W \tilde{e} + u$ (voir section 3.1);
3. d'appliquer les MCO sur $(I - \tilde{\rho} W)\, y = (I - \tilde{\rho} W)\, X\beta + u$, ce qui s'écrit aussi $\tilde{z} = \tilde{X}\beta + u$, donc $\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{z}$;
4. de reformer $\tilde{e} = y - X\tilde{\beta}$ et de retourner à 2., jusqu'à convergence.

---

[1]. D'ici la dernière section de ce chapitre, nous allons considérer sans le vérifier que l'estimation par la méthode du maximum de vraisemblance produit des estimateurs qui possèdent les propriétés habituelles désirables. A la section finale, nous chercherons à voir si pour chacun des cas présentés à la section 1), ce résultat semble raisonnable.

3.1) Estimation $\tilde{\rho}$ du maximum de vraisemblance

(2) peut s'écrire si on remplace ε par e où on a enlevé le ~ pour simplifier la notation :

$$(40) \qquad (I - \rho W)\, e = u \,,$$

$$Pe = u \,.$$

D'après (2), la vraisemblance s'écrit alors[1] :

$$\mathcal{L}(\rho, \sigma^2) = (2\pi\sigma^2)^{-N/2} |P| \exp\{-\frac{e' P'P e}{2\sigma^2}\} \,,$$

dont le log naturel s'écrit :

$$(41) \qquad L(\rho, \sigma^2) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 + \ln |P| - \frac{e' P'P e}{2\sigma^2} \,.$$

La condition du premier ordre par rapport à σ^2 conduit à $\tilde{\sigma}^2 = \frac{e' P'P e}{N}$. En substituant $\tilde{\sigma}^2$ dans (41), la fonction concentrée s'écrit :

$$(42) \qquad L^*(\rho) = -\frac{N}{2}[\ln (2\pi) + 1] - \frac{N}{2} \ln \tilde{\sigma}^2 + \ln |P| \,,$$

ce que certains écrivent souvent comme :

$$(43) \qquad L^*(\rho) = cst - \frac{N}{2} \ln [\tilde{\sigma}^2 \cdot |P|^{-2/N}] \,.$$

---

[1]. On établit ici une convention. Le déterminant d'une matrice A sera noté det A. La valeur absolue du déterminant de la matrice A sera noté $|A|$. Pour un scalaire γ, $|\gamma|$ représentera la valeur absolue de γ.

Alors $\tilde{\rho}$ est la valeur de ρ qui maximise $L^*(\rho)$. Toutefois, à chaque itération, il faut évaluer le déterminant de la matrice P. En se servant du fait que W est fixé à priori, Ord (1975) utilise l'astuce suivante :

Pour toute matrice W carrée quelconque, comme $\det [\lambda I - W] = \prod_{i=1}^{N} (\lambda - \lambda_i)$ [1], on a alors, pour P :

$$\det P = \det [I - \rho W] = \prod_{i=1}^{N} (1 - \rho \lambda_i) \quad , \tag{44}$$

où λ_i est la ième racine caractéristique de $W_{(N \times N)}$.

Maximiser (43) est donc équivalent à minimiser, par rapport à ρ, $\tilde{\sigma}^2 \cdot \left[\prod_{i=1}^{N} |(1 - \rho \lambda_i)| \right]^{-2/N}$, ou encore [2]:

$$L^*(\rho) \propto \ln \tilde{\sigma}^2 - \frac{2}{N} \sum_{i=1}^{N} \ln |(1 - \rho \lambda_i)| \quad , \tag{45}$$

où

$$\tilde{\sigma}^2 = \frac{e' P' P e}{N} \quad . \tag{45.1}$$

---

1. Voir Dhrymes (1978), p. 468, ce qui se vérifie facilement pour W symétrique. Comme on a toujours, $C' W C = \Lambda$ où $CC' = I = C'C$. Comme $\det [\lambda I - W] = \det C' \cdot \det [\lambda I - W] \det C = \det [\lambda I - \Lambda] = \prod_{i=1}^{N} (\lambda - \lambda_i)$.

2. A la section 5), nous étudierons le comportement analytique de (45) et discuterons de l'existence d'une solution qui maximise globalement cette fonction.

Lorsque W est symétrique, il existe toujours une matrice orthogonale C telle que C' W C = Λ. Pour ce cas précis de symétrie, Bodson et Peeters (1975) proposent la simplification suivante sur (45.1). Avec la transformation orthogonale e = Cz, (45.1) s'écrit :

$$\tilde{\sigma}^2 = \frac{z'C' \, [I - \rho \, C \Lambda C'] \; CC' \, [I - \rho \, C \Lambda C'] \, Cz}{N} \quad ,$$

et donc :

$$\tilde{\sigma}^2 = \frac{z' \, [I - \rho \, \Lambda]' \, [I - \rho \, \Lambda] \, z}{N} = \frac{\sum_{i=1}^{N} z_i^2 \, (1 - \rho \, \lambda_i)^2}{N} \quad . \tag{45.2}$$

De cette façon, lorsque W est symétrique, maximiser (43) est équivalent à minimiser :

$$\mathcal{L}^*(\rho) \propto \left| \prod_{i=1}^{N} (1 - \rho \, \lambda_i) \right|^{-2/N} \cdot \sum_{i=1}^{N} \frac{z_i^2 \, (1 - \rho \, \lambda_i)^2}{N} \quad . \tag{46}$$

Pour une étude qui comporte un grand nombre de régions, cette simplification dans l'évaluation de la fonction de vraisemblance peut s'avérer, à plusieurs points de vue, désirable.

A la prochaine section, on présentera l'algorithme du maximum de vraisemblance de Hepple (1976) lequel considère l'estimation jointe du vecteur $[\beta', \rho, \sigma^2]'$. On peut vérifier analytiquement l'équivalence de cet algorithme avec celui de Ord (1975).

4) Méthode du maximum de vraisemblance (Hepple (1976))

Comme le Jacobien de transformation $J(\varepsilon \to y) = 1$, le modèle SAR(1) formé par les équations (1) et (2), produit la fonction de vraisemblance suivante :

$$(47) \qquad \mathcal{L}(\beta, \sigma^2, \rho) = (2\pi \cdot \sigma^2)^{-N/2} |P| \cdot \exp\left\{-\frac{(y - X\beta)' P'P (y - X\beta)}{2\sigma^2}\right\}$$

où $P = [I - \rho W]$ et $E(\varepsilon\varepsilon') = \sigma^2 V = \sigma^2 P^{-1} P^{-1}$, car $P\varepsilon = u$, où $u \sim N(0, \sigma^2 I_N)$.

Après avoir concentré sur $\tilde{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} y$ et $\tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N}$, le problème est équivalent à minimiser seulement par rapport à ρ la fonction suivante :

$$(48) \qquad \mathcal{L}^{**}(\rho) \propto \left| \prod_{i=1}^{N} (1 - \rho \lambda_i) \right|^{-2/N} \cdot \frac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N} ,$$

dont le log naturel, similairement à (45), s'écrit :

$$(49) \qquad L^{**}(\rho) \propto \ln \tilde{\sigma}^2 - \frac{2}{N} \sum_{i=1}^{N} \ln |(1 - \rho \lambda_i)|$$

où $\tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N}$.

L'expression (48) fait ressortir tout comme dans l'estimation en série chronologique, la différence entre l'estimation du maximum de vraisemblance et celle qui minimise la forme χ^2, $\tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N}$ [1].

---

[1]. Voir Judge (1980), chapitre 5 ou Hepple (1976).

En effet, les méthodes Cochrane-Orcutt itérative et deux étapes, la méthode deux étapes de Durbin ou celle des moindres carrés non linéaires sont des méthodes qui visent à minimiser $\tilde{\sigma}^2$. Pour un AR(1) temporel, l'estimation du maximum de vraisemblance revient à minimiser :

$$(50) \qquad \mathcal{L}^{\Delta\Delta}(\rho) \propto (1-\rho^2)^{-1/N} \cdot \frac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N} \quad .$$

Comme lorsque $|\rho|$ n'est pas trop près de 1, le terme $(1-\rho^2)^{-1/N}$ devient négligeable lorsque $N \to \infty$:

$$(50.1) \qquad \lim_{N\to\infty} (1-\rho^2)^{-1/N} = 1 \quad ,$$

toutes ces procédures sont asymptotiquement équivalentes[1].

L'équivalent spatial SAR(1) de l'expression (50), comme vu à l'équation (48), s'écrit :

$$\mathcal{L}^{**}(\rho) \propto \left| \prod_{i=1}^{N} (1-\rho\,\lambda_i) \right|^{-2/N} \cdot \frac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N} \quad ,$$

où $\left| \prod_{i=1}^{N} (1-\rho\,\lambda_i) \right| = J(u \to \varepsilon)$ que l'on note J.

Pour Griffith (1980), la $\lim_{N\to\infty} J^{-2/N} = 1$. Ceci le conduit à prétendre que l'équivalence asymptotique des méthodes d'estimations AR(1) temporelles, établie à l'expression (50.1), doit aussi s'appliquer au modèle spatial SAR(1). Il existe plusieurs façons de critiquer ce point qui, on le verra, est faux. En effet :

---

[1]. Voir Judge (1982), p. 446-449.

1. contrairement à $(1-\rho^2)^{-1/N}$, où le nombre d'observations N n'est présent qu'à l'exposant, ici $J^{-2/N} = \left|\prod_{i}^{N} (1-\rho\,\lambda_i)\right|^{-2/N} = |P|^{-2/N}$. Ainsi, dans l'évaluation de la limite lorsque $N \to \infty$, il faut tenir compte du comportement de N nombres, lesquels sont élevés à l'exposant -2/N. Donc, dans l'évaluation de la $\lim_{N\to\infty} J^{-2/N}$, cet auteur n'a pas raison de ne considérer que l'exposant. De plus, lorsque N augmente, les racines λ_i se modifient;

2. $\lim_{N\to\infty} J^{-2/N}$ ne peut pas être égal à 1. En effet, on a montré précédemment qu'une méthode qui minimise $\frac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N}$ est dans le cas spatial non convergente. Afin de produire une estimation de ρ convergente, il a fallu ajouter le terme supplémentaire $J(u \to \varepsilon)$. C'est, en effet, ce qui a caractérisé les algorithmes des sections 3.1) et 4). Comme déjà mentionné, jusqu'à la fin du chapitre, on considère que la méthode du maximum de vraisemblance produit dans le cas spatial, des estimateurs convergents;

3. comme réponse à Griffith (1980), Ord (1981), par un petit exemple, montre qu'en général, $\lim_{N\to\infty} J^{-2/N} \neq 1$;

4. pour une formulation géographique rectangulaire, le modèle quadrilataire isotrope autorégressif a une structure de contiguïté telle que $\lim_{N\to\infty} J^{-2/N} = \exp\left[\sum_{j=1}^{\infty} \frac{1}{j} \binom{2j}{j}^2 \rho^{2j}\right]$ [1], ce qui est évidemment différent de 1. Ainsi, finalement, on conclut que la pénalité $J^{-2/N}$ est un terme important

---

[1]. Voir Haining (1978d), p. 497.

qu'il ne faut pas négliger. De plus, afin de produire une estimation convergente de ρ, on suggère de minimiser l'expression (48). En ce qui nous concerne, en repartant de (47), on maximisera plutôt :

$$(51) \qquad L^{**}(\rho) \propto + \ln |P| - \frac{N}{2} \ln \tilde{\sigma}^2 \quad ,$$

où $\tilde{\sigma}^2 = \dfrac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N}$.

Il est évident que maximiser (51) est équivalent à minimiser (48).

5) Etude du comportement analytique de la fonction de vraisemblance du SAR(1)

Avant d'étudier le comportement de la fonction de vraisemblance, $L^{**}(\rho) \propto + \ln |P| - \frac{N}{2} \ln \tilde{\sigma}^2$, rappelons les implications mathématiques de la stabilité spatiale et établissons certains concepts sur les trois types de matrices de contiguïté qui nous intéressent.

Le schéma autorégressif

$$(52) \qquad \varepsilon = \rho W \varepsilon + u \quad ,$$

qui s'écrit :

$$(53) \qquad (I - \rho W) \varepsilon = u$$

possède une représentation stable en terme du vecteur u lorsque :

$$(54) \qquad |\mu_{MAX} (\rho W)| < 1 \quad .$$

On a alors :

$$(55) \qquad \varepsilon = (I - \rho W)^{-1} u = \sum_{i=0}^{\infty} \rho^i W^i u \quad .$$

En terme des racines de W, il faut que [1] :

$$(56) \qquad |\rho| < \frac{1}{|\lambda_{MAX}(W)|} \quad .$$

Lorsque (56) est respecté, on définit l'équation (53) comme un processus spatial autorégressif d'ordre 1 stable.

5.1) Comportement analytique de L** pour une matrice de contiguité W quelconque

Ici, la matrice W est une matrice de contiguïté symétrique ou non, laquelle n'a subi aucun type de normalisation. On le rappelle, les matrices de contiguïté ont toujours des zéros sur la diagonale, c'est-à-dire : $tr(W) = 0$.

Selon Ord (1981), la matrice W sera toujours de signe indéterminé, c'est-à-dire qu'on aura toujours des racines caractéristiques négatives et positives. Si on les ordonne dans l'ordre croissant, on a :

$$\lambda_1, \ldots\ldots, \lambda_s, \lambda_{s+1}, \ldots\ldots, \lambda_N \quad ,$$

où λ_s est la plus grande racine négative ($|\lambda_s| \leq |\lambda_1|$) et λ_{s+1} la plus petite racine positive.

---

[1]. En effet, $\det [\rho W - \mu I_N] = 0$ s'écrit aussi $\det [W - \lambda I] = 0$ où $\rho\lambda = \mu$. Pour que la suite matricielle $[I + \rho W + \rho^2 W^2 + \ldots]$ converge, il faut que $|\mu_{MAX}(\rho W)| < 1$. Donc il faut que $|\rho\, \lambda_{MAX}| = |\rho| \cdot |\lambda_{MAX}| < 1$.

C.Q.F.D.

On peut donc écrire :

(57) $$\lambda_s^{-1} \leq \lambda_{s-1}^{-1} \leq \ldots \leq \lambda_1^{-1} < 0 < \lambda_N^{-1} \leq \ldots \leq \lambda_{s+1}^{-1}$$

où comme tr (W) = 0, $\sum_{i=1}^{N} \lambda_i = 0$.

L'expression (51) peut s'écrire grâce à (44) :

(58) $$L^{**}(\rho) \propto \sum_{i=1}^{N} \ln|(1-\rho\,\lambda_i)| - \frac{N}{2} \ln \tilde{\sigma}^2 \quad ,$$

où :

(59) $$\ln|P| = \ln\left[\prod_{i=1}^{N} |(1-\rho\,\lambda_i)|\right] = \sum_{i=1}^{N} \ln|(1-\rho\,\lambda_i)| \quad .$$

L'équation (57) définit pour la fonction de vraisemblance (58), au maximum N points d'indétermination. En effet, pour chaque valeur $\rho = \frac{1}{\lambda_i}$, on a ln (0) = $-\infty$. La solution du maximum de vraisemblance doit donc se situer sur un des segments déterminés par (57).

De plus, pour être une solution stable d'après (56), ρ doit se situer sur un des deux sous-segments suivants :

(60) $$\cdot\ \frac{1}{\lambda_1} \qquad \left]-\frac{1}{\lambda_N}, \frac{1}{\lambda_N}\right[\quad , \text{ si } |\lambda_N| > |\lambda_1| \quad ,$$

ou

(61) $$\left]\frac{1}{\lambda_1}, -\frac{1}{\lambda_1}\right[\qquad \cdot\ \frac{1}{\lambda_N} \quad , \text{ si } |\lambda_1| > |\lambda_N| \quad .$$

Ainsi, contrairement à ce qu'implique l'équation (5) de Ord (1981) pour $|\lambda_1| \neq |\lambda_N|$, l'intervalle de stabilité spatiale (suite matricielle convergente) sera toujours inférieur à celui défini par $\frac{1}{\lambda_1} < \rho < \frac{1}{\lambda_N}$.

Dans un premier temps, étudions le comportement de la fonction $\ln |P|$ par rapport à ρ. On obtient grâce aux dérivées premières et secondes :

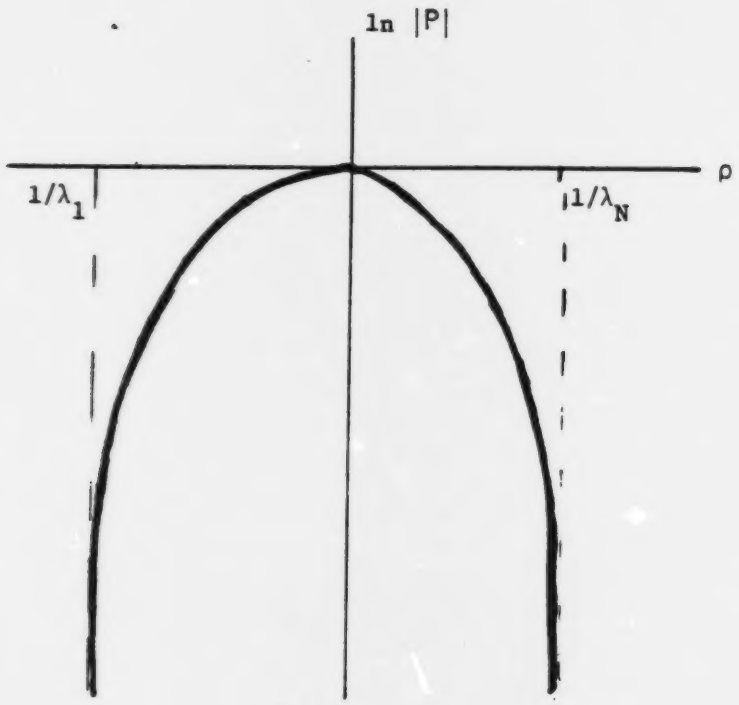
$$(62) \qquad \frac{\partial \sum_{i=1}^{N} \ln |(1-\rho\,\lambda_i)|}{\partial \rho} = - \sum_{i=1}^{N} \frac{\lambda_i}{(1-\rho\,\lambda_i)}$$

et

$$(63) \qquad \frac{\partial^2 \ln |P|}{\partial \rho^2} = - \sum_{i=1}^{N} \frac{\lambda_i^2}{(1-\rho\,\lambda_i)^2} < 0 \quad .$$

Quel que soit ρ, pourvu qu'il ne soit pas égal à un des $\frac{1}{\lambda_i}$, on a toujours $\frac{\partial^2 \ln |P|}{\partial \rho^2} < 0$. Donc, à l'intérieur de chacun des segments formés par les $\frac{1}{\lambda_i}$, on retrouve une courbe concave. Pour chaque point $\rho = 1/\lambda_i \ \forall i$, comme $|P| = \prod_{i=1}^{N} |(1-\rho\,\lambda_i)|$, alors $|P| = 0$, et alors $\ln |P| = -\infty$. On pourra se convaincre de ce résultat, au chapitre consacré à l'estimation SAR(1) du modèle qui nous intéresse. Si on se limite au segment $\frac{1}{\lambda_1} < 0 < \frac{1}{\lambda_N}$, le maximum de la forme concave se situe au point $\rho = 0$ avec $\ln |P| = 0$. En effet, lorsque $\rho = 0$, l'expression (62) est nulle par le fait que, par définition, $\sum_{i=1}^{N} \lambda_i = 0$. Si on trace cette fonction, on obtient la figure 5 suivante :

FIGURE 5



Ce résultat est conforme à l'observation 3) de Ord (1981) qui souligne que dans l'intervalle $\frac{1}{\lambda_1}$ à $\frac{1}{\lambda_N}$ lorsque ρ passe de 0 à $\frac{1}{\lambda_N}$, la dérivée première (62) passe de façon monotone de 0 à $+\infty$ et dans l'autre sens, elle passe de 0 à $-\infty$ [1].

On connaît maintenant le comportement de l'un des deux termes formant l'expression (58). Comme $-\frac{N}{2} \ln \tilde{\sigma}^2$ est une fonction qui généralement a un comportement assez lisse, ln |P| dominera et donnera à la fonction

---

[1]. Comme généralement pour une configuration irrégulière $|\lambda_s| \neq |\lambda_{s+1}|$, $|\lambda_{s-1}| \neq |\lambda_{s+2}|$ ..... $|\lambda_1| \neq |\lambda_N|$, cette fonction n'a pas à être symétrique par rapport à l'origine 0.

de vraisemblance un comportement erratique. La très forte valeur négative de ln $|P|$ à $\rho = \frac{1}{\lambda_i}\ \forall i$ et la concavité stricte de cette fonction pour chacun des segments donnera à la fonction de vraisemblance une allure similaire.

Par exemple, pour une matrice booléenne (0, 1) symétrique d'ordre 26×26 qui représente la contiguïté physique dans le modèle que l'on va estimer au chapitre V, on trace à la figure 6 la fonction de vraisemblance correspondante pour $-2 \leq \rho \leq 2$.

Pour la matrice booléenne (0, 1) 26×26, en plus des 16 racines négatives, on a les 10 racines positives suivantes. Dans l'ordre décroissant, elles sont :

$\lambda = 5,130909$	$\lambda = 1,724691$
$\lambda = 4,061591$	$\lambda = 1,302968$
$\lambda = 3,546919$	$\lambda = 0,906918$
$\lambda = 2,672019$	$\lambda = 0,700976$
$\lambda = 2,238742$	$\lambda = 0,423877$

ce qui veut dire que dans l'intervalle 0, à $\frac{1}{0,420} = 2,381$, on a les 10 points d'indétermination suivants :

$\rho = 0,194897$	$\rho = 0,579814$
$\rho = 0,246209$	$\rho = 0,767479$
$\rho = 0,281935$	$\rho = 1,102643$
$\rho = 0,374249$	$\rho = 1,426588$
$\rho = 0,446679$	$\rho = 2,359175$.

On remarque, à la figure 6, les portions de formes concaves imposées par ln $|P|$. Pour un traçage à intervalle plus petit $\Delta\rho$, les zones d'indétermination seraient plus accentuées. Cette fonction de vraisemblance est issue de la somme des fonctions ln $|P|$ et $-\frac{N}{2} \ln \tilde{\sigma}^2$ qui en se limitant à $0 \leq \rho \leq 1$ sont tracées aux figures 7 et 8 suivantes.

FIGURE 6

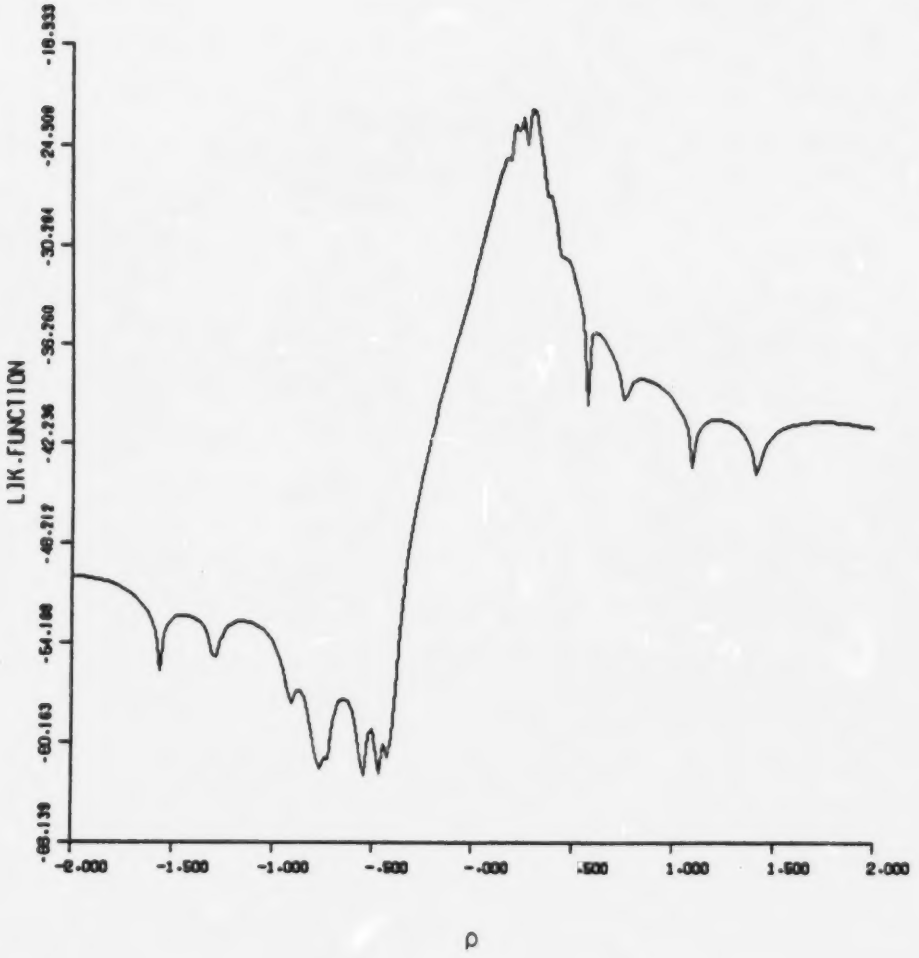

LJK-FUNCTION
-16.333
-24.509
-32.684
-40.860
-49.212
-57.563
-65.915
-74.266
-82.618
-90.969
-2.000
-1.500
-1.000
-.500
-.000
.500
1.000
1.500
2.000
ρ

FIGURE 7

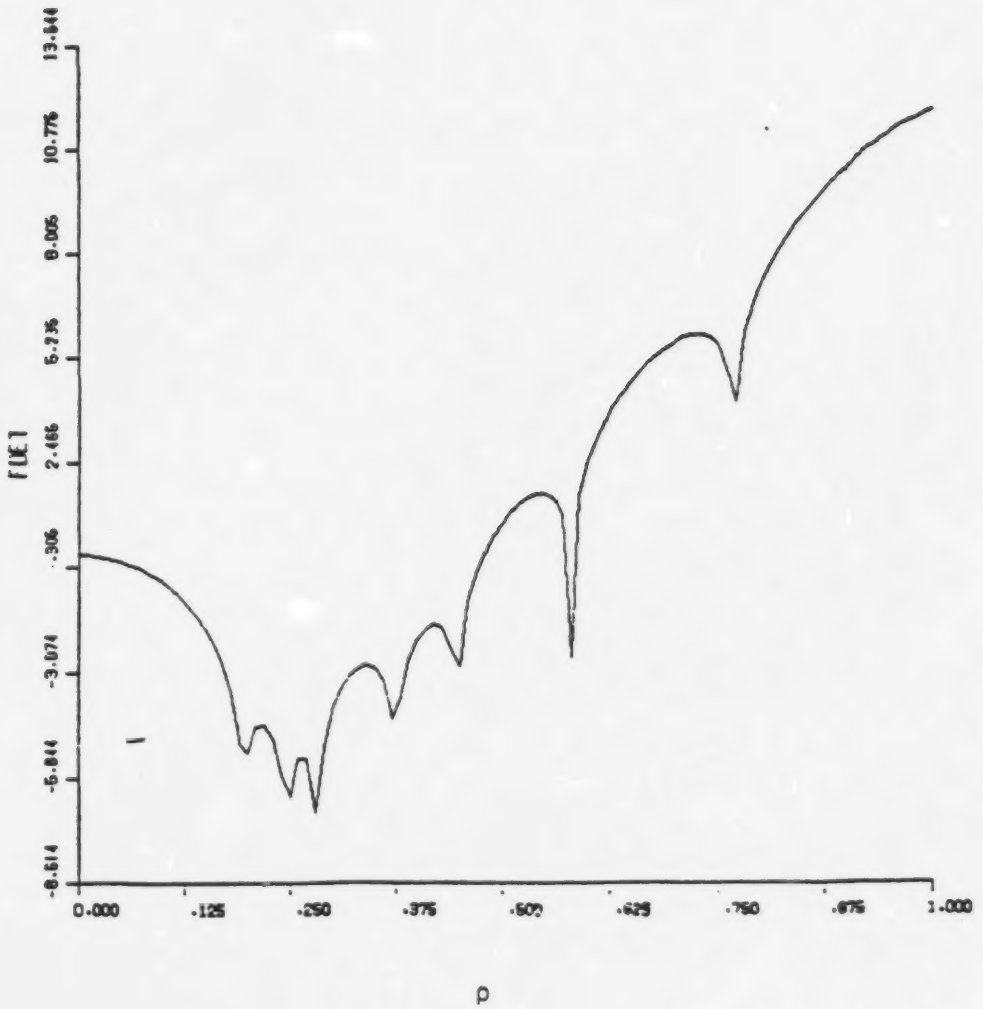
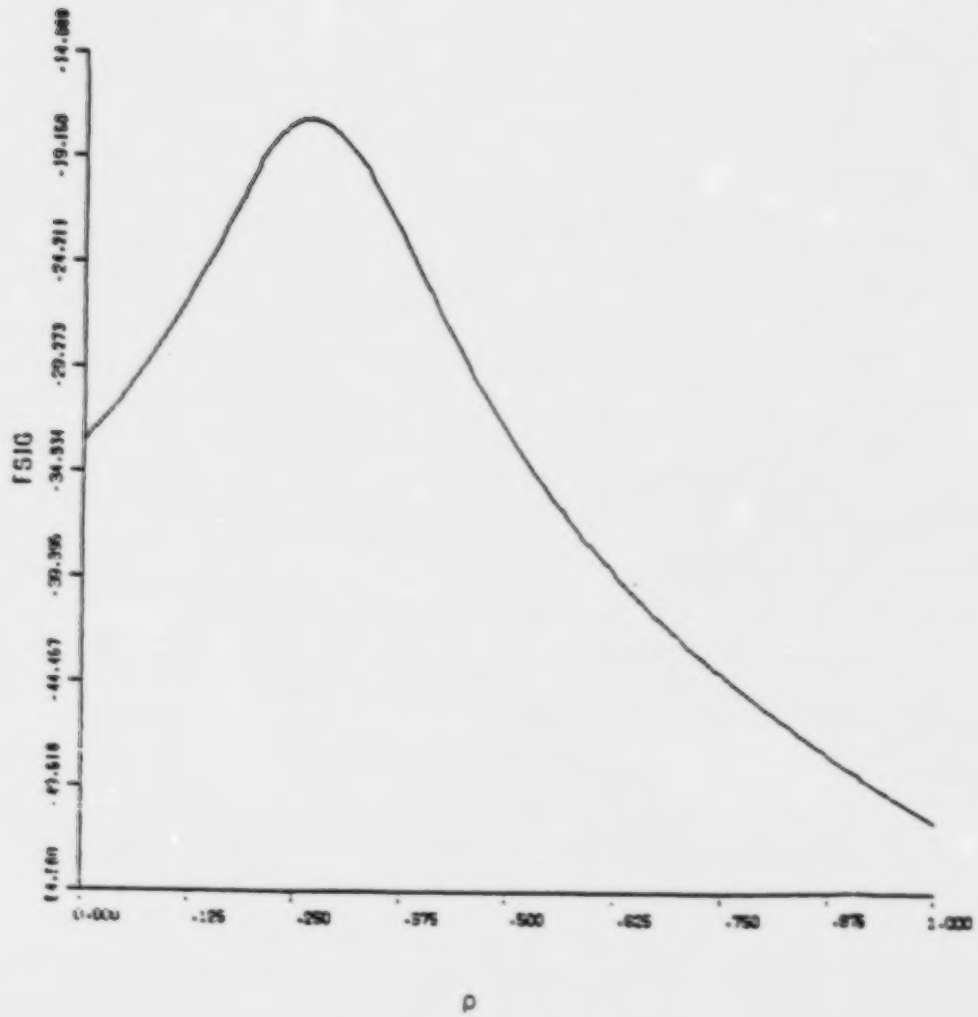

F[E]
13.644
10.776
8.005
5.236
2.466
.306
-3.074
-5.844
-8.614
0.000
.125
.250
.375
.500
.625
.750
.875
1.000
1
ρ

FIGURE 8



5.2) Comportement analytique de L** pour une matrice de contiguïté M

Ici, on veut étudier quel sera l'impact sur la fonction (51) d'imposer à la matrice de contiguïté la forme normalisée :

$$(64) \qquad m_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$$

où $m_{ii} = 0 \ \forall i$, tel que $\sum_j m_{ij} = 1 \ \forall i$. On le rappelle, ce genre de normalisation produit une matrice non symétrique. A la section 5.2.1), on montrera que $\lambda_{MAX}(M) = +1$. Ceci implique donc, d'après l'équation (56) qu'une solution stable devra se situer sur le segment $-1 < \rho < +1$, ce qui est du point de vue interprétation économique, très intéressant.

5.2.1) *Racine caractéristique dominante de la matrice de contiguïté M*

Proposition 1 : $\lambda = +1$ est toujours une solution de l'équation caractéristique

$$\det [M - \lambda I] = 0$$

Preuve

Comme $Mi = i$

$$(65) \qquad (M - 1\,I)i = 0 \ .$$

Ainsi le vecteur unitaire $\underset{(N\times 1)}{i}$ représente le vecteur propre associé à la racine caractéristique $\lambda = 1$, d'où :

$$(66) \qquad \det (M - I) = 0 \ .$$

C.Q.F.D.

On va maintenant démontrer que $\lambda = +1$ est la valeur maximale que peut prendre toute racine caractéristique de M.

Theorème 1

Si λ est une racine caractéristique de la matrice A, alors :

$$|\lambda| \leq \max_i \sum_j |a_{ij}| \tag{67}$$

Preuve

$(A - \lambda I)\, x = 0$ s'écrit : $\sum_{j=1}^{N} a_{ij}\, x_j = \lambda\, x_i, \quad i = 1,2,\ldots,N.$

$$\forall i \quad |\lambda|\,|x_i| = \Big|\sum_j a_{ij}\, x_j\Big| \leq \sum_j |a_{ij}|\,|x_j| \leq \sum_j |a_{ij}| \max_j |x_j|$$

d'où :

$$\forall i \quad |\lambda| \max_i |x_i| \leq \sum_j |a_{ij}| \cdot \max_j |x_j|$$

d'où $|\lambda| \leq \sum_j |a_{ij}|$.

C.Q.F.D.

Corollaire 1

La racine caractéristique dominante de la matrice M respecte $|\lambda| \leq 1$. De plus, $\lambda = +1$ est toujours une solution de l'équation caractéristique $\det [M - \lambda I] = 0$.

Preuve

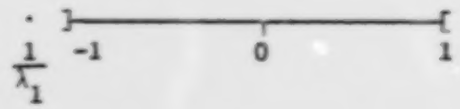
Par la proposition 1 et par le théorème 1 car $m_{ij} \geq 0$ et $\sum_j m_{ij} = 1$.

5.2.2) <u>*Comportement de la fonction* L** *dans la région stable*</u>

Ici, on applique les principes développes à la section 5.1). Soit λ_i la racine caractéristique i de M, d'après le résultat de la section 5.2.1), l'équation (57) s'écrit :

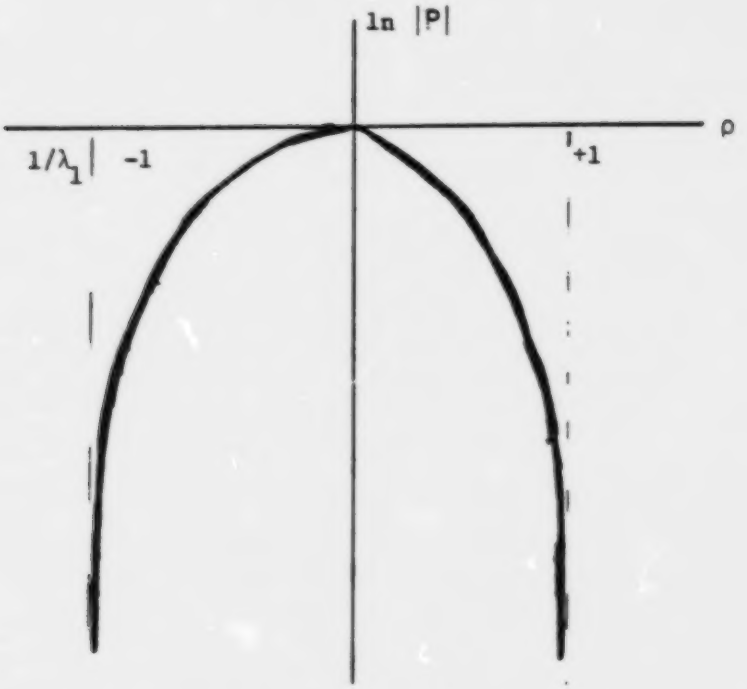
(68) $$\frac{1}{\lambda_s} \leq \frac{1}{\lambda_{s-1}} \leq \ldots \leq \frac{1}{\lambda_1} < 0 < 1 \leq \frac{1}{\lambda_{N-1}} \leq \ldots \leq \frac{1}{\lambda_{s+1}} .$$

Comme M+I, [cas où $\lambda = -1$] n'est généralement pas une matrice singulière, on aura toujours $|\lambda_1| < 1$. La solution stationnaire $-1 < \rho < 1$ se situera donc sur le sous-segment :

(69) 

Le paragraphe qui suit l'équation (63) sur le comportement de ln |P| implique donc qu'on aura comme présenté à la figure 9 :

FIGURE 9



L'étude du comportement de la fonction ln |P| pour le cas précis de la matrice M peut se faire d'une façon différente de celle présentée à la section 5.1).

En effet, comme $m_{ij} \geq 0$ $\forall ij$ et $\sum_j m_{ij} = 1$, pour $0 < \rho < 1$, $P = [I - \rho M]$ est une matrice qui possède un inverse dont tous les éléments sont positifs[1]. Ainsi, comme :

$$(70) \qquad \frac{\partial \ln |P|}{\partial \rho} = \text{tr}\left(P^{-1} \frac{\partial P}{\partial \rho}\right) = - \text{tr}\left(P^{-1} M\right) ,$$

alors $\frac{\partial \ln |P|}{\partial \rho} \leq 0$.

---

1. $(I - \rho M)^{-1} = (I + \rho M + \rho^2 M^2 + \rho^3 M^3 + \ldots)$.

En effet, comme P^{-1} est une matrice dont tous les éléments sont positifs et M une matrice avec éléments non négatifs, l'équation (70) est vérifiée pour toute valeur de ρ entre 0 et 1. Comme la dérivée seconde implique que :

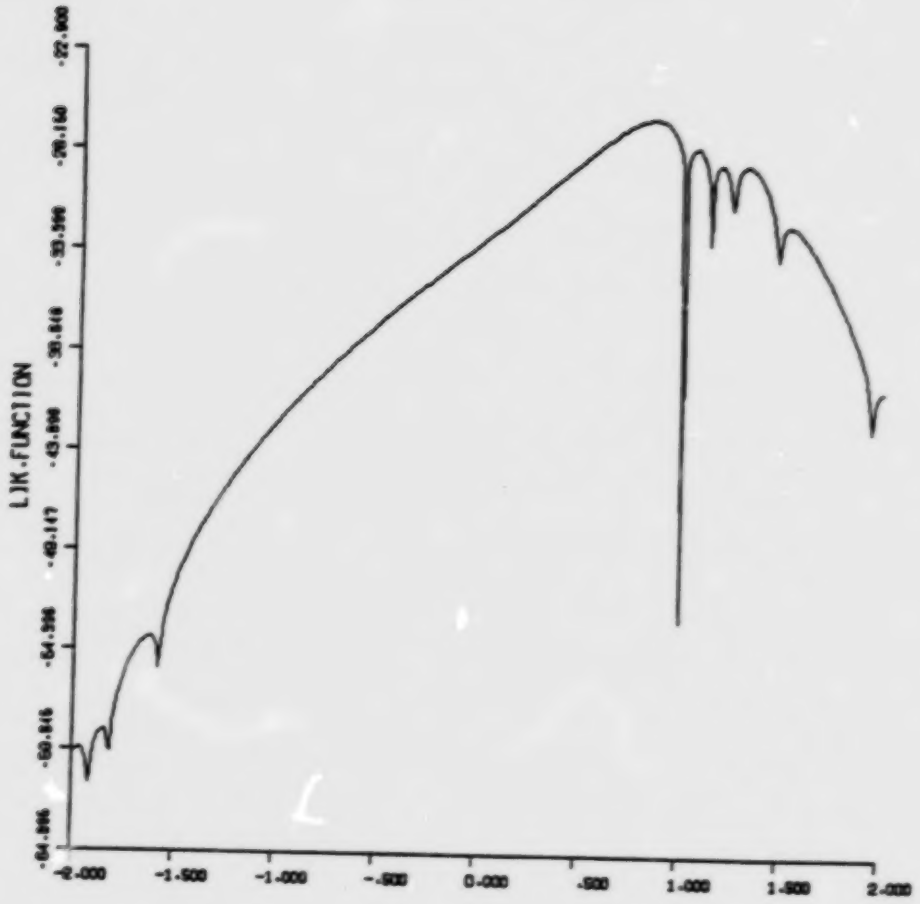
$$(71) \qquad \frac{\partial^2 \ln |P|}{\partial \rho^2} = - \text{tr} (P^{-1} M P^{-1} M) < 0 \quad , \qquad \forall \rho \in [0, 1[$$

les équations (70) et (71) confirment l'allure de la courbe dans le quadrant IV de la figure 9. Pour l'étude de la fonction dans le quadrant III, consulter Hordijk et Paelinck (1976).

Si on se limite à la région stable $-1 < \rho < 1$, dû au comportement généralement lisse de la fonction $-\frac{N}{2} \ln \tilde{\sigma}^2$, l'application de la forme concave $\ln |P|$ se reflète grandement sur la fonction de vraisemblance.

Par exemple, lorsqu'on normalise la matrice booléenne (0, 1) dont on vient de parler, on obtient la figure 10 suivante. On constate ici très bien le point d'indétermination +1 de la fonction de vraisemblance, et la forte concavité sur l'intervalle [- 1,5 et + 1] .

FIGURE 10


LIK-FUNCTION
-22.900
-28.150
-33.399
-38.648
-43.898
-49.147
-54.396
-59.645
-64.895
-2.000
-1.500
-1.000
-.500
0.000
.500
1.000
1.500
2.000
D

5.3) Comportement analytique de L** pour une matrice de contiguïté $\underset{\sim}{W}$

On montrera dans un premier temps que la matrice $\underset{\sim}{W}$ possède les mêmes racines caractéristiques que la matrice M de la section 5.2). On le rappelle, soit W une matrice symétrique, on a posé :

$$\underset{\sim}{w}_{ij} = \frac{w_{ij}}{\sqrt{\sum_i w_{ij}} \cdot \sqrt{\sum_j w_{ij}}} \quad , \tag{72}$$

et

$$m_{ij} = \frac{w_{ij}}{\sum_j w_{ij}} \quad . \tag{73}$$

Soit D une matrice diagonale avec élément $d_{ii} = \frac{1}{\sum_j w_{ij}}$, on a donc :

$$M = W \cdot D \quad , \tag{74}$$

et

$$\underset{\sim}{W} = D^{1/2} W D^{1/2} \quad . \tag{75}$$

Comme $M = D^{-1/2} \underset{\sim}{W} D^{1/2}$, les matrices M et $\underset{\sim}{W}$ sont similaires et possèdent par le fait même les mêmes racines caractéristiques[1].

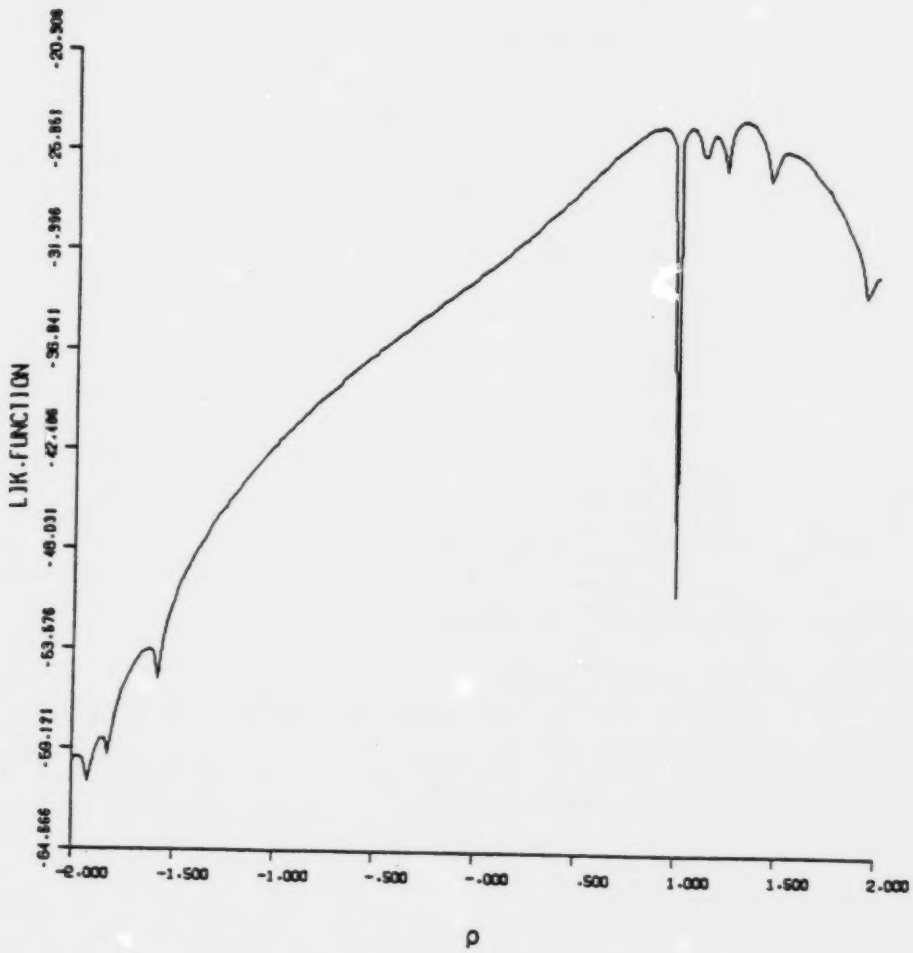
On peut ainsi conclure que la racine dominante de $\underset{\sim}{W}$ sera aussi +1. De plus, l'étude du comportement de la fonction ln |P| de la section 5.2.2) s'applique directement ici.

---

[1]. Voir Balestra (1972), p. 140.

Entre les deux approches, la différence essentielle réside dans le fait que $[I - \rho M] \neq [I - \rho W]$. La fonction de vraisemblance aura donc un comportement différent de celui impliqué par la matrice M. Pour se convaincre, en appliquant la normalisation w_{ij} à notre matrice booléenne (0, 1) symétrique, on obtient la figure 11. On peut alors remarquer le comportement quelque peu différent de celui de la figure 10. Comme attendu, les points d'indétermination sur les deux figures sont les mêmes.

FIGURE 11



On remarque, grâce aux figures 6 et 11, la présence d'une solution non stable supérieure à la solution sur le domaine $\frac{-1}{|\lambda_{MAX}|} < \rho < \frac{1}{|\lambda_{MAX}|}$. Toutefois, à la figure 10, ce phénomène ne se produit pas.

On se demande alors si c'est le fait d'imposer la symétrie qui crée cette solution maximale en dehors de la région stable, ou bien est-ce imputable uniquement au modèle qui nous intéresse ?

Il serait alors très intéressant de vérifier ce phénomène dans le cas d'autres modèles. Pour l'instant, on ne peut pas s'expliquer pourquoi.

A la prochaine section, on donnera un bref compte-rendu critique sur les expériences empiriques qui utilisent une formulation SAR(1) résiduelle.

6) Quelques résultats empiriques concernant l'estimation du SAR(1) résiduel

Le SAR(1) résiduel s'écrit :

(76) $$y = X\beta + \varepsilon \quad ,$$

(77) $$\varepsilon = \rho W \varepsilon + u \quad , \qquad u \sim N(0, \sigma^2 I) \quad .$$

Jusqu'à maintenant, les éléments qui forment la matrice de contiguïté spatiale ont toujours été considérés fixes, c'est-à-dire qu'on postule généralement pour les w_{ij}, une forme booléenne (0, 1) ou une forme fonctionnelle dont les paramètres sont fixés à priori. Afin de garantir la stabilité spatiale sur le domaine $|\rho| < 1$, les auteurs utilisent fréquemment la matrice de contiguïté normalisée non symétrique M. Comme l'estimation MCO de ρ, à l'équation (77), produit un estimateur $\hat{\rho}$ non convergent, la majorité des expériences empiriques visent la minimisation de :

$$(78) \qquad \mathcal{L}^{**}(\rho) \propto \{ \prod_{i=1}^{N} |(1 - \rho \lambda_i)| \}^{-2/N} \cdot \frac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N} \quad ,$$

ou du logarithme naturel de $\mathcal{L}^{**}(\rho)$:

$$(79) \qquad L^{**}(\rho) \propto - \frac{2}{N} \sum_{i=1}^{N} \ln |(1 - \rho \lambda_i)| + \ln \tilde{\sigma}^2 \quad ,$$

où $\tilde{\sigma}^2 = \dfrac{(y - X\tilde{\beta})' V^{-1} (y - X\tilde{\beta})}{N}$.

Comme technique de minimisation, la plupart procèdent à une procédure de balayage sur la fonction (79)[1]. A la section 6.1), on présentera des procédures que l'on a classifiées comme non convergentes, car elles visent la minimisation de $\tilde{\sigma}^2$. En 6.2), on s'attardera aux résultats empiriques de méthodes qui maximisent la vraisemblance de l'échantillon.

6.1) Méthodes d'estimation inspirées de la série chronologique

Hordijk (1974) propose d'appliquer la méthode itérative de Cochrane-Orcutt et la méthode en 2 étapes de Durbin (1960) sur le processus spatial résiduel d'ordre k suivant :

$$(80) \qquad \varepsilon = \rho_1 C(1) \varepsilon + \rho_2 C(2) \varepsilon + \ldots + \rho_k C(k) \varepsilon + u \quad ,$$

où C(g) est la matrice de contiguïté physique booléenne d'ordre g.

Comme (80) peut s'écrire :

$$(81) \qquad \varepsilon = C^* \underset{(k \times 1)}{\rho} + u \quad ,$$

où $C^* = [C(1) \varepsilon \vdots C(2) \varepsilon \vdots \ldots \vdots C(k) \varepsilon]$; on a alors :

---

[1]. A notre connaissance, seul Haining (1978b), dans une application du modèle isotrope Q.A. du chapitre I, a utilisé une méthode d'estimation itérative avec dérivées.

$$(82) \qquad \underset{(k\times 1)}{\hat{\rho}} = [C^{*\prime}\ C^{*}]^{-1}\ C^{*\prime}\ \varepsilon \quad .$$

L'estimation, dans les deux méthodes, étant basée sur des applications successives des MCO sur (76) et (81), entraîne, comme on l'a montré, des estimateurs non convergents.

L'application de ce modèle vise à étudier l'impact qu'a produit l'introduction, vers la fin des années 1950, de la télévision, sur le niveau d'assistance dans les salles de cinéma. Cet auteur a comparé le changement entre 1957 et 1958 de ces variables pour 48 régions des Pays-Bas. Comme il le souligne, la méthode itérative n'a pas convergé même après un nombre imposant d'itérations. Il a donc abandonné cette approche. Les résultats les plus concluants et intéressants de la procédure en 2 étapes de Durbin ont été produits par la version avec schéma autorégressif résiduel d'ordre 1.

Hordijk (1979) estime par la méthode itérative de Cochrane-Orcutt un modèle SAR(1) qui cherche à expliquer l'augmentation entre 1962 et 1970 de l'emploi industriel dans 40 des régions des Pays-Bas. Pour des matrices de contiguïté W booléenne (0, 1) normalisée et non normalisée, il obtient $\hat{\rho} \simeq - 0{,}64$. Les résultats de la section 5) précédente nous pousse fortement à croire que si Hodijk avait tenu compte du facteur $[\prod^{N}|(1-\rho\ \lambda_i)|]^{-2/N}$ dans son estimation, il aurait obtenu, pour une matrice W non normalisée, une valeur $\tilde{\rho}$ passablement différente de celle obtenue avec une matrice W normalisée.

Bodson et Peeters (1975), pour expliquer la demande de travail dans les industries qui produisent des biens et services voués à l'exportation, proposent, par une méthode des moindres carrés non linéaires, de minimiser u'u de l'équation (77).

6.2) Résultats d'estimation du maximum de vraisemblance pour le SAR(1) résiduel [1]

Hepple (1976) applique le SAR(1) résiduel avec matrice de contiguïté booléenne normalisée M à un modèle qui cherche à évaluer les effets que produisent, sur l'achat d'automobiles aux Etats-Unis, les différences régionales dans les taxes de vente et les charges de transport imputables à l'achat de voitures neuves. A la lumière des t de Student, l'estimation du modèle MCO par Hanna (1966) a donné de bons résultats.

Hepple minimise (78) par la méthode itérative de Powell sans dérivées. Les conclusions empiriques de l'étude montrent qu'après correction de l'autocorrélation spatiale, le modèle postulé par Hanna (1966) perd de son intérêt.

Hordijk et Paelinck (1976) proposent une méthode d'estimation du maximum de vraisemblance pour un modèle autorégressif spatial avec schéma :

$$(83) \qquad \varepsilon = [\rho\, C(1) + \rho^2\, C(2) + \ldots \rho^k\, C(k)]\varepsilon + u \ .$$

---

[1]. Pour une approche bayesienne, consulter Anselin (1982) et Hepple (1979).

Pour k = 1, et $\rho > 0$, leur méthode suggère d'appliquer une procédure de balayage sur la région stable $0 \leq \rho < \frac{1}{\lambda_{MAX}[C(1)]}$. Pour un problème d'explication des différences régionales dans le revenu de 40 régions des Pays-Bas, leur méthode a produit des résultats intéressants. En effet, par rapport à l'estimation MCO, deux coefficients ont changé de signe dans la bonne direction.

Ord (1975) a appliqué la procédure Cochrane-Orcutt modifiée de la section 2)[1] au modèle qui nous intéressera aux chapitres V à VII. On ne présentera alors pas ici un compte-rendu de son expérience empirique. Il en est de même pour les estimations du maximum de vraisemblance de ce même modèle par Cliff et Ord (1981), Burridge (1981), Blommestein (1983) et Bivand (1984). En effet, on aura plus loin, l'occasion de revenir sur les résultats empiriques de ces auteurs.

On peut maintenant formuler une critique commune à toutes les approches que l'on vient d'examiner. Dans la formulation du schéma autorégressif spatial, on considère toujours connus les éléments de la matrice de contiguïté. De façon générale, pour une matrice de contiguïté W dont les éléments w_{ij} s'écrivent :

$$(84) \qquad w_{ij} = \begin{cases} g\,(V_{ij}^1,\ldots,V_{ij}^P\ ;\ \theta_1,\ldots,\theta_H) & \text{, si i et j sont contigus} \\ 0 & \text{, sinon ,} \end{cases}$$

où on a P variables pertinentes et H paramètres inconnus, toutes ces procédures fixent à priori le vecteur $\theta_{H\times 1} = (\theta_1 \ \ldots \ \theta_H)'$, où par la suite, lorsque désiré, on applique une normalisation quelconque.

---

1. Pour une application de cette méthode dans le domaine de l'agriculture, consulter Haining (1978d).

Afin de combler cette lacune, on proposera au prochain chapitre une méthode d'estimation du maximum de vraisemblance pour les paramètres de la forme (84) et les paramètres (β, σ^2 et ρ) des équations (76) et (77). Une application sera présentée aux chapitres V à VII.

7) Sur les propriétés asymptotiques des estimateurs du maximum de vraisemblance pour un modèle SAR(1) résiduel

L'étude des propriétés asymptotiques des estimateurs du maximum de vraisemblance pour un modèle avec régresseurs fixes dont les erreurs sont i.i.d. peut se faire avec l'aide des récents travaux de Burguete, Gallant et Souza (1982), Gallant et Holly (1980), Gourieroux, Monfort et Trognon (1984) et White (1982). Concernant la convergence des estimateurs du maximum de vraisemblance, lorsque $\lim_{N\to\infty} \frac{1}{N} L_N(\cdot) \xrightarrow{P} \bar{L} < \infty$, où L_N représente le logarithme naturel de la fonction de vraisemblance, si $\bar{L}$ possède un maximum unique, les estimateurs du maximum de vraisemblance sont convergents[1]. White et Domowitz (1983) étendent ces résultats aux cas où les observations y_i sont dépendantes entres elles. Leur étude étant spécifique au cas où la dépendance entre les y_i ne se fait que par l'arrière, [classe des modèles ARMA temporels], les outils développés dans cet article sont inapplicables ici.

Ayant besoin de résultats, nous basons notre étude sur l'article de Magnus (1978). Ce dernier établit les conditions qui garantissent, pour

---

1. Comme Gallant et Holly le soulignent, l'approche que prennent ces différents auteurs permet d'établir, pour l'étude des propriétés des estimateurs, des conditions qui sont relativement faciles à vérifier.

les estimateurs du maximum de vraisemblance d'un modèle MCG avec matrice de covariances des résidus estimée, la convergence, la normalité et l'efficacité asymptotique. Comme cet article concerne une formulation plus générale que celle des articles déjà cités, ces dernières conditions sont, dans la pratique, plus difficiles à vérifier.

A la section 7.1), nous présentons les résultats principaux de Magnus (1978). En 7.2), nous appliquons la méthodologie de cet auteur au modèle SAR(1) résiduel.

7.1) Résultats de Magnus (1978) concernant les propriétés des estimateurs (MCG) du maximum de vraisemblance

Avant d'établir les conditions désirées, on doit énoncer un certain nombre d'hypothèses. Nous avons le modèle de régression linéaire suivant :

(85) $\quad y = X\beta + \varepsilon$

avec

H1 : $\varepsilon \sim N$.

H2 : $E(\varepsilon) = 0 \quad E(\varepsilon\varepsilon') = \Omega$ où $\Omega = \Omega(\theta)$ est une matrice définie positive dont les éléments $\omega(\theta)$ sont des fonctions du vecteur de paramètre $\theta = (\theta_1, \theta_2, \ldots, \theta_m)'$. Les $\omega(\theta)$ sont des fonctions continues différentiables jusqu'au second ordre par rapport aux $\theta_i \quad i = 1, \ldots, m$, $\theta_i \in \Theta$.

H3 : X est une matrice fixe de rang K.

H4 : Il n'y a pas de paramètres communs aux vecteurs β et θ.

Comme la fonction de vraisemblance de l'échantillon s'écrit :

$$(86) \qquad \mathcal{L}(\beta,\theta|y, X) = (2\pi)^{-N/2} \, |\Omega|^{-1/2} \exp\{-\frac{\varepsilon' \, \Omega^{-1} \, \varepsilon}{2}\} \ ,$$

Magnus montre que sous H1 à H4, la matrice d'information $\psi = - E[H]$ (H représentant la Hessienne) est :

$$(87) \qquad \psi = \begin{bmatrix} X' \, \Omega^{-1} \, X & 0 \\ 0 & \frac{1}{2} \, \psi_\theta \end{bmatrix}$$

où

$$(88) \qquad (\psi_\theta)_{ij} = \text{tr} \left(\frac{\partial \Omega^{-1}}{\partial \theta_i} \, \Omega \, \frac{\partial \Omega^{-1}}{\partial \theta_j} \, \Omega\right) \qquad i,j = 1,\ldots,m \ .$$

H5 : Pour s'assurer que ψ_θ est non singulière, on doit faire l'hypothèse que les θ_i $i = 1,\ldots,m$, sont identifiables.

Selon cette hypothèse, à chaque vecteur $\theta^1 \neq \theta^2$, il correspond une matrice $\Omega^1 \neq \Omega^2$.

Pour faire l'étude des propriétés asymptotiques des estimateurs du maximum de vraisemblance, Magnus ajoute[1] :

H6 : $\lim_{N\to\infty} \frac{1}{N} \psi_{ij}$ est une fonction finie de β et θ, $\forall \ \beta \in B$ et $\forall \ \theta \in \Theta$

$(i,j = 1,\ldots,K+m)$.

---

[1]. Lorsque les θ_i sont identifiables, l'hypothèse cruciale est que $\Omega(\theta)$ doit être définie positive. Sous ces conditions, on peut vérifier que ψ_θ est nécessairement définie positive. Il en est de même pour ψ. (Voir Magnus (1978), p. 288). Les conditions H6 et H7 sont des conditions suffisantes pour que $\text{plim} \ (- h_{ij}) = \psi_{ij}$.

H7 : $\lim_{N\to\infty} \text{var}\left(\frac{1}{N} h_{ij}\right)$ converge vers zero

$\forall\ \beta \in B$ et $\theta \in \Theta$ $\quad (i,j = 1,\ldots,K+m)$

Théorème

Les estimateurs du maximum de vraisemblance $\hat{\beta}$ et $\hat{\theta}$ de β et θ du modèle (85) sont, sous les hypothèses H1 à H7, convergents, possèdent une loi limite normale et sont asymptotiquement efficaces

Preuve : Magnus (1978), p. 295.

7.2) Résultats de Magnus (1978) pour un modèle SAR(1) résiduel

Selon notre notation pour le modèle SAR(1) résiduel, nous avons :

$$y = X\beta + \varepsilon \quad ,$$

(89)

$$\varepsilon = \rho W \varepsilon + u \quad , \qquad u \sim N(0, \sigma^2 I) \quad .$$

Grâce à l'équation (47), soit $P = I - \rho W$, $V(\varepsilon) = \sigma^2 V = \sigma^2 P^{-1} P^{-1}$, avec $V^{-1} = P'P$.

Comme V est fonction de ρ et comme σ^2 est un paramètre à estimer, on peut noter comme Magnus (1978) $V(\varepsilon) = \Omega(\theta) = \sigma^2 V(\rho)$, où $\theta = [\sigma^2\ \rho]'$.

Vérifions maintenant ce qu'implique pour le SAR(1) résiduel, les hypothèses de la section 7.1).

H2 : D'après cette hypothèse, $V^{-1} = P'P$ doit être une matrice définie positive pour toute valeur de ρ appartenant à son domaine paramétrique.

Résultat 1

Lorsque le domaine paramétrique de ρ est $-\infty \leq \rho \leq \infty$, l'estimation du maximum de vraisemblance de ρ est non convergente.

Preuve

Pour que la procédure du maximum de vraisemblance soit convergente, les hypothèses H1 à H7 doivent être respectées.

Comme d'après l'équation (44) $\det P = \prod_{i=1}^{N} (1 - \rho\lambda_i)$ où λ_i est la ième racine caractéristique de W, et comme $\det V^{-1} = \det P \cdot \det P$, dès que ρ peut être égal à $\frac{1}{\lambda_i}$, la matrice des variances et des covariances des résidus est non définie positive ce qui viole l'hypothèse H2.

Une autre façon d'obtenir ce résultat consiste à écrire :

$$V^{-1} = P'P = (I - \rho W)' (I - \rho W) \tag{90}$$

$$= I - 2\rho W + \rho^2 W'W \quad .$$

Comme W peut s'écrire :

$$W = C \Lambda C' \tag{91}$$

avec $C'C = CC' = I$, on a :

$$V^{-1} = C [I - 2\rho \Lambda + \rho^2 \Lambda^2] C' \quad . \tag{92}$$

Ainsi, dès que $\rho = 1/\lambda_i$, V^{-1} devient non définie positive.

Ce résultat, allié à ceux établis au chapitre III pousse à conclure qu'en absence de constrainte sur le domaine de ρ, l'estimation du maximum de vraisemblance pour les cas ④, ⑤ et ⑥ sera non convergente.

Remarques

1) Comme dans les cas ①, ② et ③, on se limite à la région stationnaire $-1 < \rho < 1$ de ρ, la violation de l'hypothèse H2 pour ces modèles spatiaux ne risque pas de se produire. En effet, on peut vérifier que pour ces trois modèles, lorsque ρ passe de -1 à +1, V demeure définie positive.

2) Dans la même ligne d'idée, comme lorsqu'on applique à la matrice W la normalisation M ou $\underset{\sim}{W}$, on a nécessairement (68) et comme λ_{MIN} est nécessairement ≥ -1, le domaine stable $-1 < \rho < 1$ est tel que toute valeur de ρ sur ce domaine donne une matrice V définie positive.

3) Comme dans le cas ③ $\lambda_i = 0 \quad \forall i = 1,\ldots,N$, l'estimation du maximum de vraisemblance des paramètres de ce modèle est équivalente à une estimation des moindres carrés non linéaires, car $|J|$ est alors égal à 1 dans la fonction de vraisemblance.

En considérant que pour le modèle SAR(1) résiduel considéré, les hypothèses H3 à H5 sont respectées, il reste à vérifier que sont respectées les hypothèses H6 et H7. Pour appliquer la méthodologie de Magnus (1978), il suffit d'appliquer les concepts d'espérance mathématique et de variance aux éléments formés par les équations (19) à (24) du chapitre IV. En consultant ces équations, on s'aperçoit rapidement de la difficulté de vérifier ces con-

ditions dans la pratique. Malgré cela, il ressort tout de même le fait que $V(\varepsilon)$ doit être une matrice définie positive. A cause de cela, nous obtenons :

Résultat 2

Lorsqu'on se limite aux versions SAR(1) résiduelles avec matrice de contiguïté normalisée M ou $\underset{\sim}{W}$, si la valeur vraie $\rho_o \in]-1, +1]$, la procédure du maximum de vraisemblance risque d'être convergente.

Remarque

Comme souligné à la fin de la section 6), au chapitre IV, on présente une procédure d'estimation du maximum de vraisemblance pour les paramètres du modèle SAR(1) résiduel où les éléments de :

$$w_{ij} = g\,(v_{ij}^1, v_{ij}^2, \ldots, v_{ij}^P, \theta_1, \ldots, \theta_H) \tag{93}$$

sont des fonctions estimables. Les résultats des chapitres III et V à VII poussent à croire que les conclusions précédentes demeurent valables pour l'ensemble des paramètres $[\beta' \theta' \rho \sigma^2]'$.

CHAPITRE IV

ESTIMATION D'UN PROCESSUS AR(2) RÉSIDUEL SPATIAL AUTORÉGRESSIF

On veut estimer le modèle suivant pour R régions[1] :

$$(1) \qquad \underset{(R\times 1)}{y} = \underset{(R\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(R\times 1)}{\varepsilon} \quad ,$$

$$(2) \qquad \underset{(R\times 1)}{\varepsilon} = \rho_1\ \underset{(R\times R)}{B}\ \underset{(R\times 1)}{\varepsilon} + \rho_2\ \underset{(R\times R)}{C}\ .\ \underset{(R\times 1)}{\varepsilon} + \underset{(R\times 1)}{u} \quad ,$$

où $u \sim N(0, \sigma^2 I_R)$, avec $B = B(\theta)$, la matrice R×R de contiguïté d'ordre 1 et $C = C(\theta)$, la matrice R×R de contiguïté d'ordre 2, où B et C ont des éléments b_{ij} et c_{ij} fonctions d'un certain vecteur de paramètres $\underset{(H\times 1)}{\theta}$, c'est-à-dire $\theta' = (\theta_1, \theta_2, \ldots, \theta_H)$.

1) Fonction de vraisemblance

L'équation (2) peut s'écrire :

$$(3) \qquad [I_R - \rho_1 B - \rho_2 C]\ \varepsilon = u \quad ,$$

$$(4) \qquad P(\rho_1, \rho_2, \theta)\ \varepsilon = u \quad .$$

Pour simplifier la notation, on dira :

$$(5) \qquad P\varepsilon = u \ , \quad \text{et } \varepsilon = P^{-1} u \ .$$

Ainsi, le jacobien de transformation s'écrit :

$$(6) \qquad J(u \to \varepsilon) = |P| \quad ,$$

---

[1]. On développe ici la solution d'un SAR(2) résiduel car, comme on pourra le constater lors de chapitres ultérieurs, cette formulation est directement compatible avec celle d'un schéma autorégressif appliqué à des données de flux. Par souci de comptabilité de notation entre ces deux types de formulation, on considérera ici que l'on possède R régions.
L'algorithme d'estimation du SAR(1) qui nous intéresse ici est un cas particulier du SAR(2). En effet, il suffit d'imposer $\rho_2 = 0$ pour avoir la solution d'un problème avec autocorrélation résiduelle d'ordre 1. Les conditions de stabilité et les propriétés théoriques du SAR(2) pourront être étudiées lors de papiers subséquents.

c'est donc la valeur absolue du déterminant de P. Comme matrice variance-covariance pour ε, on a :

$$(7) \qquad V(\varepsilon) = \sigma_u^2 \, P^{-1} \, P^{-1\prime} = \sigma_u^2 \, V \quad .$$

Posons,

$$(8) \qquad \gamma' = (\rho_1, \rho_2, \theta_1, \theta_2, \ldots, \theta_H) \quad ,$$

et $\sigma_u^2 = \sigma^2$.

Alors la fonction de vraisemblance de l'échantillon s'écrit :

$$(9) \qquad \mathcal{L}(\beta, \sigma^2, \gamma; y) = (2\pi \cdot \sigma^2)^{-R/2} \, |P| \, \exp\left\{- \frac{(y - X\beta)' \, V^{-1} \, (y - X\beta)}{2\sigma^2}\right\} \quad ,$$

dont le logarithme naturel s'écrit :

$$(10) \qquad L(\beta, \sigma^2, \gamma; y) = - \frac{R}{2} \ln (2\pi) - \frac{R}{2} \ln \sigma^2 + \ln |P| - \frac{(y - X\beta)' \, V^{-1} \, (y - X\beta)}{2\sigma^2} \quad ,$$

où, on le rappelle, on a :

$P = P(\gamma)$ avec $V = V(\gamma) = P^{-1} \, P^{-1\prime}$ et donc $V^{-1} = P'P$.

On maximise (10) avec l'aide de l'algorithme de minimisation de Davidon (1959) et Fletcher et Powell (1963), lequel requiert l'évaluation des dérivées premières.

2) Dérivées premières

$$(11) \qquad \underset{(K\times 1)}{\frac{\partial L}{\partial \beta}} = + \frac{X' \, V^{-1} \, (y - X\beta)}{\sigma^2} \quad ,$$

$$(12) \qquad \frac{\partial L}{\partial \sigma^2} = -\frac{R}{2\sigma^2} + \frac{(y - X\beta)' V^{-1} (y - X\beta)}{2\sigma^4} \quad ,$$

$$(13) \qquad \frac{\partial L}{\partial \gamma_i} = \text{tr} \left(P^{-1} \frac{\partial P}{\partial \gamma_i}\right) - \frac{(y - X\beta)'}{2\sigma^2} \left[P' \frac{\partial P}{\partial \gamma_i} + \frac{\partial P'}{\partial \gamma_i} P\right] (y - X\beta)$$

$$= \text{tr} \left(P^{-1} \frac{\partial P}{\partial \gamma_i}\right) - \frac{(y - X\beta)'}{\sigma^2} P' \frac{\partial P}{\partial \gamma_i} (y - X\beta) \quad , \qquad i = 1,\ldots,H+2.$$

Egalant à zéro les équations (11) et (12), on obtient :

$$(14) \qquad \beta^* = (X' V^{-1} X)^{-1} X' V^{-1} y \quad ,$$

et

$$(15) \qquad \sigma^{*2} = \frac{(y - X\beta^*)' V^{-1} (y - X\beta^*)}{R} \quad .$$

β^* et σ^{*2} sont des fonctions explicites du vecteur γ qui, pour toutes valeurs de γ, rendent les équations (11) et (12) nulles. Ceci nous permet de concentrer (10) par rapport à β et σ^2 ce qui ramène le problème à une maximisation par rapport aux γ_i seulement. En effet, on a :

$$(16) \qquad L^{**} (\gamma; y, \beta^*, \sigma^{*2}) = -\frac{R}{2} \ln [(2\pi) + 1] - \frac{R}{2} \ln \sigma^{*2} + \ln |P| \quad ,$$

avec, selon (13) :

$$(17) \qquad \frac{\partial L^{**}}{\partial \gamma_i} = \text{tr} \left(P^{-1} \frac{\partial P}{\partial \gamma_i}\right) - \frac{1}{\sigma^{*2}} (y - X\beta^*)' P' \frac{\partial P}{\partial \gamma_i} (y - X\beta^*) \quad ,$$

qui doit être égal à zéro $\forall \gamma_i$, $i = 1,\ldots,H+2$, pour que (16) soit maximale. Evalué à ce point, on obtient l'estimation du maximum de vraisemblance $\hat{\gamma}$, $\hat{\sigma}^2 = \frac{(y - X\hat{\beta})' \hat{V}^{-1} (y - X\hat{\beta})}{R}$ et $\hat{\beta} = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} y$.

Afin d'évaluer l'équation (17), on doit connaître $\frac{\partial P}{\partial \gamma_i}$. On a :

(18) $\gamma_i =$

	$\frac{\partial P}{\partial \gamma_i}$
ρ_1	$- B$
ρ_2	$- C$
θ_h	$-[\rho_1 \frac{\partial B}{\partial \theta_h} + \rho_2 \frac{\partial C}{\partial \theta_h}]$

où $P = [I_R - \rho_1 B - \rho_2 C]$ et où $h = 1,\ldots,H$.

3) Dérivées secondes

(19) $$\frac{\partial^2 L}{\partial \beta \partial \beta'} = - \frac{X' V^{-1} X}{\sigma^2} ,$$

(20) $$\frac{\partial^2 L}{\partial \beta \partial \sigma^2} = - \frac{X' V^{-1} \varepsilon}{\sigma^4} ,$$

(21) $$\frac{\partial^2 L}{\partial \beta \partial \gamma_i} = \frac{X'}{\sigma^2} [P' \frac{\partial P}{\partial \gamma_i} + \frac{\partial P'}{\partial \gamma_i} P] \varepsilon ,$$

(22) $$\frac{\partial^2 L}{\partial \sigma^2 \partial \sigma^2} = \frac{R}{2\sigma^4} - \frac{\varepsilon' V^{-1} \varepsilon}{\sigma^6} ,$$

$$(23)\qquad \frac{\partial^2 L}{\partial\sigma^2\partial\gamma_i} = \frac{\varepsilon'\,[P'\,\frac{\partial P}{\partial\gamma_i} + \frac{\partial P'}{\partial\gamma_i}\,P]\,\varepsilon}{2\sigma^4} = \frac{\varepsilon'\,P'\,\frac{\partial P}{\partial\gamma_i}\,\varepsilon}{\sigma^4}\ ,$$

$$= \frac{1}{\sigma^4}\,u'\,\frac{\partial P}{\partial\gamma_i}\,\varepsilon\quad .$$

Donc, on a, grâce à (18) :

$$(23.1)\qquad \gamma_i = \begin{array}{c|c} & \dfrac{\partial^2 L}{\partial\sigma^2\partial\gamma_i} \\ \hline \rho_1 & -\dfrac{u'\,B\,\varepsilon}{\sigma^4} \\ \rho_2 & -\dfrac{u'\,C\,\varepsilon}{\sigma^4} \\ \theta_h & \dfrac{-\,u'\,[\rho_1\,\frac{\partial B}{\partial\theta_h} + \rho_2\,\frac{\partial C}{\partial\theta_h}]\,\varepsilon}{\sigma^4} \end{array}\ ,$$

$$(24)\qquad \frac{\partial^2 L}{\partial\gamma_i\partial\gamma_j} = \mathrm{tr}\,[-\,P^{-1}\,\frac{\partial P}{\partial\gamma_j}\,P^{-1}\,\frac{\partial P}{\partial\gamma_i} + P^{-1}\,\frac{\partial^2 P}{\partial\gamma_i\partial\gamma_j}]$$

$$-\,\frac{\varepsilon'}{\sigma^2}\,[\frac{\partial P'}{\partial\gamma_j}\,\frac{\partial P}{\partial\gamma_i} + P'\,\frac{\partial^2 P}{\partial\gamma_i\partial\gamma_j}]\,\varepsilon\quad ,$$

où, pour h et g = 1,...,H, on a, grâce à (18) :

$$\gamma_j =$$

$\frac{\partial^2 p}{\partial\gamma_i \partial\gamma_j}$	ρ_1	ρ_2	θ_g
ρ_1	0	0	$-\frac{\partial B}{\partial \theta_g}$
ρ_2	0	0	$-\frac{\partial C}{\partial \theta_g}$
θ_h	$-\frac{\partial B}{\partial \theta_h}$	$-\frac{\partial C}{\partial \theta_h}$	$-[\rho_1 \frac{\partial^2 B}{\partial \theta_h \partial \theta_g} + \rho_2 \frac{\partial^2 C}{\partial \theta_h \partial \theta_g}]$

(24.1) $\gamma_i =$

4) Evaluation de la matrice variance-covariance

Comme par les équations (20) et (21) $E\left[\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\right] = 0$ et $E\left[\frac{\partial^2 L}{\partial\beta\partial\gamma_i}\right] = 0$, on vérifie que $\hat{\beta}$ est asymptotiquement indépendant de $\hat{\sigma}^2$ et $\hat{\gamma}$. Grâce à la partition de la matrice d'information, la matrice variance-covariance de $\hat{\beta}$ se calcule par la négative de l'inverse numérique de la matrice de l'équation (19). Pour $\hat{\sigma}^2$ et $\hat{\gamma}$, on se limite à la sous-matrice formée par les équations (22) à (24). Ainsi, on trouve :

$$V(\hat{\beta}) = \hat{\sigma}^2 \, (X' \, \hat{V}^{-1} \, X)^{-1}$$

$$V\left(\begin{bmatrix}\hat{\sigma}^2 \\ \hat{\gamma}\end{bmatrix}\right) = -\left[\frac{\partial^2 L}{\partial \begin{bmatrix}\sigma^2 \\ \gamma\end{bmatrix} \partial[\sigma^2 \, \gamma']}\right]^{-1}\Bigg|_{\substack{\gamma = \hat{\gamma} \\ \sigma^2 = \hat{\sigma}^2}} .$$

5) Formules programmées pour les dérivées secondes

En notant $D_h = \rho_1 \frac{\partial B}{\partial \theta_h} + \rho_2 \frac{\partial C}{\partial \theta_h}$, $D_g = \rho_1 \frac{\partial B}{\partial \theta_g} + \rho_2 \frac{\partial C}{\partial \theta_g}$ et

$D_{hg} = \rho_1 \frac{\partial^2 B}{\partial \theta_h \partial \theta_g} + \rho_2 \frac{\partial^2 C}{\partial \theta_h \partial \theta_g}$, la formule (24) s'écrit :

$$(25) \qquad \frac{\partial^2 L}{\partial \rho_1^2} = - \operatorname{tr} (P^{-1} B P^{-1} B) - \frac{\varepsilon' B'B \varepsilon}{\sigma^2} \quad .$$

$$(26) \qquad \frac{\partial^2 L}{\partial \rho_1 \partial \rho_2} = - \operatorname{tr} (P^{-1} C P^{-1} B) - \frac{\varepsilon' C'B \varepsilon}{\sigma^2} = \frac{\partial^2 L}{\partial \rho_2 \partial \rho_1} \quad ,$$

$$(27) \qquad \frac{\partial^2 L}{\partial \rho_2^2} = - \operatorname{tr} (P^{-1} C P^{-1} C) - \frac{\varepsilon' C'C \varepsilon}{\sigma^2} \quad .$$

$$(28) \qquad \frac{\partial^2 L}{\partial \theta_h \partial \rho_1} = - \operatorname{tr} (P^{-1} B P^{-1} D_h) - \operatorname{tr} (P^{-1} \frac{\partial B}{\partial \theta_h}) - \frac{\varepsilon' B' D_h \varepsilon}{\sigma^2} + \frac{u' \frac{\partial B}{\partial \theta_h} \varepsilon}{\sigma^2} \quad ,$$

$$(29) \qquad \frac{\partial^2 L}{\partial \theta_h \partial \rho_2} = - \operatorname{tr} (P^{-1} C P^{-1} D_h) - \operatorname{tr} (P^{-1} \frac{\partial C}{\partial \theta_h}) - \frac{\varepsilon' C' D_h \varepsilon}{\sigma^2} + \frac{u' \frac{\partial C}{\partial \theta_h} \varepsilon}{\sigma^2} \quad ,$$

$$(30) \qquad \frac{\partial^2 L}{\partial \theta_h \partial \theta_g} = - \operatorname{tr} (P^{-1} D_g P^{-1} D_h) - \operatorname{tr} (P^{-1} D_{hg}) - \frac{\varepsilon' D'_g D_h \varepsilon}{\sigma^2} + \frac{u' D_{hg} \varepsilon}{\sigma^2} \quad .$$

6) Modification de l'algorithme suite à l'imposition de la normalisation symétrique

Dans cette section, on présente les modifications qu'implique la normalisation symétrique, dans la formulation de l'algorithme du SAR(2), c'est-à-dire ce que l'on doit modifier dans l'algorithme lorsqu'on passe de B et C à $\underset{\sim}{B}$ et $\underset{\sim}{C}$ [1].

Afin de ne pas allonger inutilement le texte, on limitera notre exercice au passage de B vers $\underset{\sim}{B}$. Pour $C \rightarrow \underset{\sim}{C}$, la démarche est identique. Soit B la matrice de contiguïté d'ordre 1, la matrice de contiguïté normalisée symétrique $\underset{\sim}{B}$ peut s'écrire :

$$\underset{\sim}{B} = D\,B\,D' \quad , \tag{31}$$

où $D = D'$ est une matrice diagonale telle que :

$$D = \begin{bmatrix} \left(\sum_{j=1}^{R} b_{1j}\right)^{-1/2} & 0 & \cdots & 0 \\ 0 & \left(\sum_{j=1}^{R} b_{2j}\right)^{-1/2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\sum_{j=1}^{R} b_{Rj}\right)^{-1/2} \end{bmatrix} . \tag{32}$$

L'équation (31) est telle que terme à terme, on a :

$$\underset{\sim}{b}_{ij} = \frac{b_{ij}}{\sqrt{\sum_{i=1}^{R} b_{ij}} \cdot \sqrt{\sum_{j=1}^{R} b_{ij}}} \quad . \tag{33}$$

---

[1]. Par exemple, afin que l'algorithme SAR(1) puisse produire une estimation tant pour un modèle avec matrice de contiguïté symétrique non normalisée B que pour un modèle avec matrice de contiguïté symétrique normalisée $\underset{\sim}{B}$, on présente ici les modifications algébriques qu'il faut apporter à l'algorithme.

6.1) Dérivées premières

$$(34) \qquad \frac{\partial B}{\partial \theta_h} = \frac{\partial D}{\partial \theta_h} \cdot B \cdot D' + D \frac{\partial B}{\partial \theta_h} D' + D B \frac{\partial D'}{\partial \theta_h} \quad ,$$

où $\frac{\partial d_{ij}}{\partial \theta_h} = 0 \quad \forall i \neq j$ et

$$(35) \qquad \frac{\partial d_{ii}}{\partial \theta_h} = \frac{\partial \left(\sum_{j=1}^{R} b_{ij} \right)^{-1/2}}{\partial \theta_h} = -\frac{1}{2} \left(\sum_{j=1}^{R} b_{ij} \right)^{-3/2} \sum_{j=1}^{R} \frac{\partial b_{ij}}{\partial \theta_h}$$

$$= -\frac{d_{ii}}{2 \sum_{j=1}^{R} b_{ij}} \cdot \sum_{j=1}^{R} \frac{\partial b_{ij}}{\partial \theta_h} \quad .$$

En posant :

$$(36) \qquad h_{ii} = -\frac{\sum_{j=1}^{R} \frac{\partial b_{ij}}{\partial \theta_h}}{2 \sum_{j=1}^{R} b_{ij}} \quad ,$$

on a donc :

$$(37) \qquad \frac{\partial d_{ii}}{\partial \theta_h} = h_{ii} \cdot d_{ii} \quad .$$

Faisant de même pour chaque élément d_{ii} de D, on a finalement, matriciellement :

$$(38) \qquad \frac{\partial D}{\partial \theta_h} = \frac{\partial D'}{\partial \theta_h} = H \cdot D = D' H' \quad ,$$

où D et H sont des matrices diagonales.

Si on applique (38) à l'équation (34), on obtient :

$\frac{\partial \tilde{B}}{\partial \theta_h} = H\,D\,B\,D' + D\,\frac{\partial B}{\partial \theta_h}\,D' + D\,B\,D'\,H'$, ce qui s'écrit :

$$(39) \qquad \frac{\partial \tilde{B}}{\partial \theta_h} = H\,\underset{\sim}{B} + D\,\frac{\partial B}{\partial \theta_h}\,D' + \underset{\sim}{B}\,H' \quad [1] .$$

6.2) Dérivées secondes

Soit $\frac{\partial D}{\partial \theta_h} = H \cdot D = D'\,H'$, on aura :

$$(40) \qquad \frac{\partial D}{\partial \theta_\ell} = L\,D = D'\,L' \quad ,$$

car $\frac{\partial d_{ii}}{\partial \theta_\ell} = -\frac{\sum_{j=1}^{R} \frac{\partial b_{ij}}{\partial \theta_\ell}}{2 \sum_{j=1}^{R} b_{ij}} \cdot d_{ii}$, ce que l'on définit :

$$= \ell_{ii} \cdot d_{ii} \quad .$$

De (39), on obtient :

$$(41) \qquad \frac{d^2 \tilde{B}}{\partial \theta_h \partial \theta_\ell} = \frac{\partial H}{\partial \theta_\ell} \cdot \underset{\sim}{B} + H\,\frac{\partial \tilde{B}}{\partial \theta_\ell} + \underset{\sim}{B}\,\frac{\partial H'}{\partial \theta_\ell} + \frac{\partial \tilde{B}}{\partial \theta_\ell}\,H'$$

$$+ L\,D\,\frac{\partial B}{\partial \theta_h}\,D' + D\,\frac{\partial^2 B}{\partial \theta_h \partial \theta_\ell}\,D' + D\,\frac{\partial B}{\partial \theta_h}\,D'\,L' \quad .$$

---

[1]. Elément par élément, pour la programmation, on écrit :

$$\frac{\partial \underset{\sim}{b_{ij}}}{\partial \theta_h} = h_{ii}\,\underset{\sim}{b_{ij}} + d_{ii}\,\frac{\partial b_{ij}}{\partial \theta_h}\,d_{jj} + \underset{\sim}{b_{ij}}\,h_{jj} = \underset{\sim}{b_{ij}}\,(h_{ii} + h_{jj}) + d_{ii}\,\frac{\partial b_{ij}}{\partial \theta_h}\,d_{jj} .$$

Si on regroupe les termes, on a :

$$(42) \qquad \frac{\partial^2 \underset{\sim}{B}}{\partial\theta_h \partial\theta_\ell} = \frac{\partial H}{\partial\theta_\ell}\underset{\sim}{B} + \underset{\sim}{B}\frac{\partial H'}{\partial\theta_\ell} + H\frac{\partial \underset{\sim}{B}}{\partial\theta_\ell} + \frac{\partial \underset{\sim}{B}}{\partial\theta_\ell}H' + L\underset{\sim}{\frac{\partial B}{\partial\theta_h}} + \underset{\sim}{\frac{\partial B}{\partial\theta_h}}L' + \underset{\sim}{\frac{\partial^2 B}{\partial\theta_h\partial\theta_\ell}} \quad [1],$$

où, par convention, pour toute matrice A différente de B, avec l'aide de D, la matrice calculée en (32), on écrit $\underset{\sim}{A} = DAD'$.

Il nous reste à connaître $\frac{\partial H}{\partial\theta_\ell}$. Comme $h_{ii} = -\dfrac{\sum_{j=1}^{R}\frac{\partial b_{ij}}{\partial\theta_h}}{2\sum_{j=1}^{R} b_{ij}}$,

$$(43) \qquad \frac{\partial h_{ii}}{\partial\theta_\ell} = -\frac{1}{2\sum_{j=1}^{R} b_{ij}}\left[\sum_{j=1}^{R}\frac{\partial^2 b_{ij}}{\partial\theta_h\partial\theta_\ell} + 2\,h_{ii}\sum_{j=1}^{R}\frac{\partial b_{ij}}{\partial\theta_\ell}\right],$$

donc, matriciellement, on a :

$$(44) \qquad \frac{\partial H}{\partial\theta_\ell} = K + 2HL$$

---

[1]. On peut vérifier l'exactitude de ce résultat par une dérivation terme à terme à partir de la note précédente. En effet, on obtient, pour la programmation :

$$\underset{\sim}{\frac{\partial^2 b_{ij}}{\partial\theta_h\partial\theta_\ell}} = \underset{\sim}{b_{ij}}\left(\frac{\partial h_{ii}}{\partial\theta_\ell} + \frac{\partial h_{jj}}{\partial\theta_\ell}\right) + \underset{\sim}{\frac{\partial b_{ij}}{\partial\theta_\ell}}\left(h_{ii} + h_{jj}\right) + \underset{\sim}{\frac{\partial b_{ij}}{\partial\theta_h}}\left(\ell_{ii} + \ell_{jj}\right) + \underset{\sim}{\frac{\partial^2 b_{ij}}{\partial\theta_h\partial\theta_\ell}}.$$

avec $$k_{ii} = - \frac{\sum_{j=1}^{R} \frac{\partial^2 b_{ij}}{\partial\theta_h \partial\theta_\ell}}{2 \sum_{j=1}^{R} b_{ij}}$$

$$h_{ii} = - \frac{\sum_{j=1}^{R} \frac{\partial b_{ij}}{\partial\theta_h}}{2 \sum_{j=1}^{R} b_{ij}} \quad \text{et} \quad \ell_{ii} - \frac{\sum_{j=1}^{R} \frac{\partial b_{ij}}{\partial\theta_\ell}}{2 \sum_{j=1}^{R} b_{ij}} \quad .$$

Pour la programmation, on aura :

(45) $$\frac{\partial h_{ii}}{\partial\theta_\ell} = k_{ii} + 2\, h_{ii}\, \ell_{ii} \quad .$$

CHAPITRE V

UNE APPLICATION : L'ACCESSIBILITÉ ET LA STRUCTURE SPATIALE DE L'ÉCONOMIE DE LA RÉPUBLIQUE D'IRLANDE

Dans son article publié en 1968, "Accessibility and the Spatial Structure of the Irish Economy", P.M. O'Sullivan veut tester l'existence d'une relation fonctionnelle entre la configuration géographique de l'économie de la république d'Irlande en 1960 et la structure spatiale des différents réseaux de transports existants à ce moment. Pour ce faire, il régresse plusieurs indicateurs d'intensité de l'activité économique par comté sur différentes mesures d'accessibilité. Grâce à la théorie des graphes, pour caractériser de façon empirique l'infrastructure d'un réseau de transport R, O'Sullivan définit pour chacun des 26 comtés de l'étude, une mesure relative d'accessibilité A(i,R)[1]. Parmi les différentes mesures ainsi construites, O'Sullivan retient principalement la mesure d'accessibilité par route ARA (arterial road accessibility). Il justifie sa préférence sur le fait qu'en Irlande, la majorité des déplacements des passagers et des marchandises se font par route.

> "One problem which arises is which network to measure accessibility on. ... Attention was concentrated on the arterial road network proposed in 1964 as a result of a traffic census. This consists of the major links in the currently designated trunk road system. ... Railways carry a very small proportion of freight and passenger movements in Ireland and so position on the rail network may be deemed not so important as position on the road network. In 1960, 79 per cent of freight ton mileage in the Republic was road-borne and in 1964 83 per cent. Thus road is becoming more dominant. In 1961, 93 per cent of passenger mileage was on road."

Au niveau agrégé, pour représenter la variation géographique de l'activité économique, O'Sullivan utilise le revenu personnel per capita par comté, ce qui comprend les revenus de tous les secteurs de l'économie exception

---

[1]. L'accessibilité d'un comté i est donné par : $A(i,R) = \sum_{j=1}^{N} d_{ij}, \quad i = 1,\ldots,N$

où d_{ij} est une métrique de distance, de temps ou de coût de déplacement pour le réseau R.

faite des paiements de transfert, des revenus de pension et d'immigrants. Sur la base de résultats significatifs selon les tests t, les comtés les plus accessibles[1] (la partie sud-est de l'Irlande[2]) possèdent un revenu per capita plus élevé. En effet, en régressant le revenu personnel per capita sur l'accessibilité par route, O'Sullivan explique 54% de la variabilité géographique du revenu per capita par comté.

Dans la décomposition sectorielle de son étude, O'Sullivan considère finalement l'agriculture. En 1961, le secteur agricole tient une place importante, car, en effet, il constitue le plus grand employeur. Ajoutons d'ailleurs qu'à ce moment, 60% des exportations de l'Irlande sont d'origine agricole. O'Sullivan note alors une dualité dans l'économie agricole de l'Irlande. Selon lui, elle s'explique par la structure géographique et par l'emplacement des principaux marchés domestiques et ports d'exportations. De fait, dans les zones moins accessibles (nord-ouest), caractérisées par des fermes de petite taille, on retrouve une agriculture de subsistance alors que la culture dans la région sud-est est plutôt orientée vers la commercialisation locale et vers l'exportation. Pour réfléter de façon empirique ce phénomène, cet auteur utilise la mesure de subsistance suivante : le pourcentage de l'output total agricole destiné à la consommation sur la ferme[3]. (On-farm consumption of produce as a percentage of gross agricultural output per county.)

---

[1]. C'est-à-dire possédant les plus petits ARA.

[2]. Voir la figure 2.

[3]. Dorénavant, on dira : pourcentage de la production autoconsommée.

Selon nos attentes, il doit y avoir une relation positive entre cette mesure et l'accessibilité par route. En effet, on s'attend à ce que les endroits moins accessibles consacrent une partie plus importante de leur production agricole à l'autoconsommation sur la ferme. Les résultats d'O'Sullivan, pour cette régression, sont significatifs et les signes respectent les attentes. D'ailleurs, l'ARA explique à 70% la variabilité dans cette mesure de subsistance.

Résultats de la régression[1]

y = le pourcentage de l'output agricole autoconsommé,

X = arterial road accessibility (ARA).

(1) $y = \underset{(-2,7538)}{-8,4478} + \underset{(7,7263)}{0,00527}\, X$

$R^2 = 0,6966$ $\qquad \hat{\sigma}^2 = 12,5373$ $\qquad \ln \mathcal{L} = -32,3872$.

Les chiffres entre parenthèses représenteront toujours les t de Student. De plus, dans la présentation de résultats empiriques, ln $\mathcal{L}$ désignera la valeur numérique du logarithme naturel de la fonction concentrée.

Les figures 1 et 2 suivantes confirment nos attentes concernant la pente de la droite de la régression. Comme y représente ici un pourcentage (y contraint au domaine [0,1]), une procédure plus appropriée que l'application des MCO au modèle linéaire $y_t = \alpha + \beta X_t + \varepsilon_t$ aurait été de formuler un modèle logit agrégé de la forme $\ln (\frac{y_t}{1 - y_t}) = \alpha + \beta X_t + v_t$. Malgré cela, dans un objectif de comparaison de résultats avec ceux d'autres recherches

---

[1]. Avec l'aide des données fournies à la table 8.1 de la monographie de Cliff et Ord (1981), on a pu retrouver les résultats de cette dernière expérience. Voir appendice I.

nous étudierons le problème de l'autocorrélation spatiale pour le modèle linéaire $y_t = \alpha + \beta X_t + \varepsilon_t$ défini à l'équation (1).

FIGURE 1

Variable dépendante

Δ 10%
o 10% - 20%
□ 20%

FIGURE 2

ARA

Δ 30 - 40
o 40 - 50
□ 50 - +

En étudiant le signe des résidus calculés (figure 3), Cliff et Ord ont remarqué une très forte autocorrélation résiduelle. En effet, pour toutes les spécifications de matrice de contiguïté W d'ordre 1 utilisées dans leur test d'autocorrélation spatiale résiduelle I_{MCO}^{CO}, ils ont dû rejeter l'hypothèse d'absence d'autocorrélation spatiale. La valeur numérique positive de I_{MCO}^{CO} et le regroupement spatial des erreurs calculées de même signe sont des indicateurs de la présence d'un problème d'autocorrélation spatiale résiduelle positive.

FIGURE 3

Ce qu'il faut surtout retenir de cette estimation, c'est la répartition spatiale des résidus calculés. A cause de cette structure spatiale résiduelle si remarquable, de nombreux chercheurs dans le domaine spatial se sont récemment attardés aux problèmes économétriques de cette régression. Ce sont principalement l'analyse de spécification et l'estimation en présence d'autocorrélation spatiale qui ont retenu l'attention.

Afin de produire des résidus calculés spatialement non correlés, Cliff et Ord (1981), section 8.6, dans un effort de spécification, ont ajouté un certain nombre de variables explicatives. En général, bien que ces auteurs aient été capables de s'expliquer les régions de surestimation (résidus négatifs) et de sous-estimation (résidus positifs), leurs tentatives du côté empirique se sont avérées vaines. Par exemple, ils remarquent que, globalement, les résidus négatifs correspondent aux zones les mieux reliées par rail avec Dublin (zone ombragée). Comme ils le soulignent, Dublin est le marché principal d'écoulement de la production agricole du pays. En ajoutant au modèle différentes variables pour représenter ce fait, ils ont été incapables de casser, de façon satisfaisante, l'autocorrélation spatiale résiduelle[1]. En considérant alors la spécification d'origine (1) comme correcte, ces auteurs proposent par la suite d'estimer ce modèle selon une formulation spatiale autorégressive résiduelle d'ordre 1. Ils estiment alors, à la section 9.4 de leur livre, le modèle suivant, qui s'écrit selon leur notation :

---

[1]. Pour d'autres exemples, consulter cette référence. Suite à cette phase de spécification, ces auteurs concluent que la configuration spatiale résiduelle de la figure 3 n'est pas due à un problème de variables manquantes. Pour notre part, les différentes tentatives de spécification ont aussi produit des résultats non concluants.

$$(2) \qquad y = \alpha + \beta X + \varepsilon \quad ,$$

$$(3) \qquad \varepsilon = \rho W \varepsilon + u \quad ,$$

où $u \sim N(0, \sigma^2 I_N)$, W est une matrice de contiguïté normalisée telle que $\sum_{j=1}^{N} w_{ij} = 1 \ \forall i$, avec $w_{ij} = \frac{w^*_{ij}}{\sum_{j=1}^{N} w^*_{ij}}$, où $w^*_{ij} = d_{ij}^{-\theta_1} \beta_{i(j)}^{\theta_2} \ \forall i \neq j$, et où θ_1 et θ_2 sont à priori fixés à 1. d_{ij} représente la distance entre i et j et $\beta_{i(j)}$, la proportion de périmètre de contact entre les zones i et j.

De cette façon, les seuls éléments w_{ij} non nuls sont ceux des zones contiguës. L'estimation en déviation par rapport à la moyenne conduit aux résultats suivants[1] :

$$(4) \qquad y - \bar{y} = \underset{(4,2857)}{0,003} (X - \bar{X})$$

$$R^2 = 0,86 \ , \quad \hat{\rho} = \underset{(5,2)}{0,78} \ , \quad \hat{\sigma}^2 = 5,73 \ , \quad \text{log likelihood = non disponible .}$$

Dans une application spatiale de l'analyse du facteur commun (common factor analysis), Burridge (1981), Blommestein (1983) et Bivand (1984) catégorisent les équations (2) et (3) comme un modèle contraint du modèle régressif/autorégressif spatial suivant :

$$(5) \qquad y = \tilde{\alpha} + \beta X + \gamma W X + \rho W y + u \quad ,$$

---

[1]. Les résultats que l'on présente ici sont en fait ceux de Burridge (1981). En effet, tel que mentionné dans son papier, dû à une erreur dans l'algorithme de Cliff et Ord, les résultats de cette régression présentée à la section 9.4 de leur monographie, sont erronés. Sauf pour l'écart-type de $\hat{\rho}$, les résultats de Blommestein (1983) coïncident généralement avec ceux de Burridge (1981).

car si on impose la contrainte $\gamma = -\rho\beta$ (hypothèse de facteur commun), l'équation (5) se réécrit comme suit :

$$(6) \qquad (I - \rho W)\, y = \bar{\alpha} + \beta\, (I - \rho W)\, X + u \quad .$$

En posant $\bar{\alpha}i = \alpha\, (I - \rho W)i = \alpha\, (1 - \rho)i$ (où i est le vecteur unité), lorsque W est une matrice normalisée ($\sum_{j=1}^{N} w_{ij} = 1 \quad \forall i$), l'équation (6) n'est rien d'autre qu'une réécriture des équations (2) et (3). En effet, selon ces dernières :

$$(7) \qquad (y - \alpha - \beta X) = (I - \rho W)^{-1}\, u = \varepsilon \quad .$$

Selon l'évidence statistique, ces auteurs avec $\alpha = ,05$ doivent rejeter l'hypothèse $\gamma + \beta\rho = 0$. Selon leurs estimations, le modèle (5) s'avère plus vraisemblable que le modèle (6) et concluent que Cliff et Ord ont introduit dans le modèle (5) une restriction qui se révèle fausse. Burridge avoue toutefois qu'après estimation, les résidus du modèle (5) paraissent encore autocorrélés. A cause de cela et suite à la grande popularité du SAR(1), on se limitera tout de même exclusivement à l'étude du modèle (6).

Le souci de comparaison de nos résultats avec ceux des derniers auteurs mentionnés, le grand intérêt que suscite cette régression et la persistance de l'autocorrélation spatiale dans les résidus de ce modèle sont tous des facteurs qui nous incitent à appliquer notre procédure à ce phénomène. Alors, dans ce qui va suivre, en considérant comme correcte la spécification d'O'Sullivan (1968), on cherchera à estimer les paramètres des fonctions w_{ij} qui forment le processus autorégressif d'ordre 1 résiduel.

CHAPITRE VI

ÉTUDE ET IMPLICATION

DE DIFFÉRENTES FORMES FONCTIONNELLES BINAIRES POUR LES w_{ij}

Dans les chapitres précédents, on a introduit trois formes de matrices de contiguïté. La matrice W est une matrice de contiguïté symétrique ou non qui n'a subi aucun type de normalisation. La matrice M est une matrice de contiguïté normalisée sur les lignes, c'est-à-dire que $\sum_{i=1}^{N} m_{ij} = 1 \ \forall i$. On a finalement proposé la matrice de contiguïté normalisée symétrique que l'on a noté $\underset{\sim}{W}$. Dans les prochaines sections, en plus de rappeler les grandes lignes concernant ces matrices, on va étudier les implications théoriques et empiriques de chacune de ces formulations.

Dans l'estimation en présence d'autocorrélation spatiale, plusieurs se limitent souvent à la forme booléenne suivante[1] :

$$(1) \qquad w_{ij} \begin{cases} = 1 & , \quad \text{si } i \text{ et } j \text{ sont contiguës} , \\ = 0 & , \quad \text{ailleurs} . \end{cases}$$

Afin d'assurer la propriété intéressante de stabilité spatiale sur le domaine $|\rho| < 1$, ces auteurs estiment généralement leur modèle avec la version normalisée de (1), c'est-à-dire :

$$(2) \qquad m_{ij} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}} \qquad \forall i \neq j ,$$

tel que $\sum_{j=1}^{N} m_{ij} = 1 \quad \forall i$.

---

1. Voir Bivand (1984), Hepple (1976) et Brandsma et Ketellapper (1979a) pour une version normalisée m_{ij} et Hordijk (1979) pour une application avec w_{ij} booleen non normalisée.

Dans un premier temps, on aimerait ici inspecter les implications de la forme fonctionnelle (1) pour, par la suite, s'attarder à la forme normalisée (2).

1) La forme binaire (0, 1) standard (non normalisée)

Selon les équations (2) et (3) du chapitre précédent, notre modèle autorégressif spatial résiduel d'ordre 1, SAR(1) s'écrit :

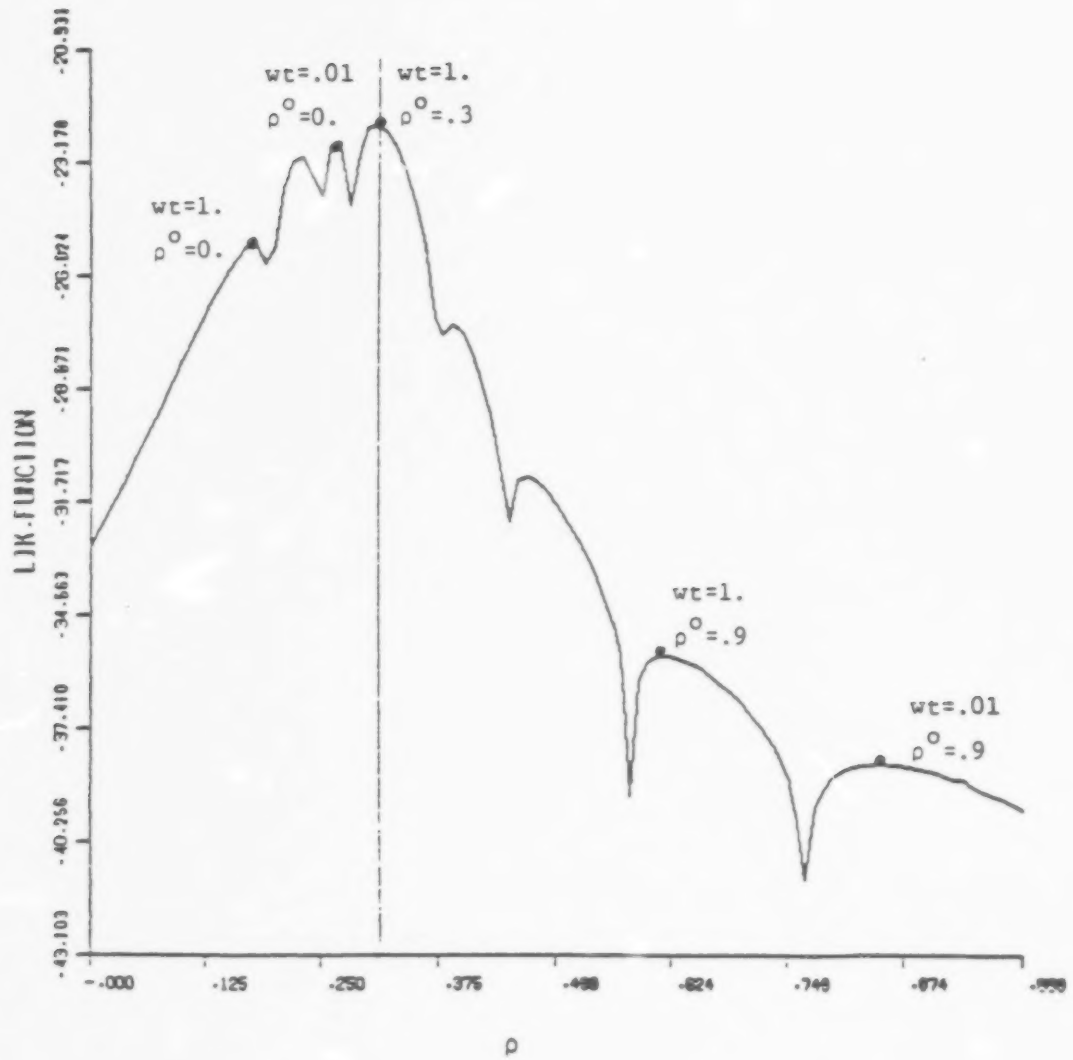
$$y = \alpha + \beta X + \varepsilon \quad ,$$

$$\varepsilon = \rho W \varepsilon + u \quad ,$$

où ρ représente le coefficient d'autocorrélation spatiale résiduelle et où, dans le cas qui nous concerne, $w_{ij} = 1$, si i et j sont contiguës et $w_{ij} = 0$ ailleurs. On le rappelle, X est l'indice d'accessibilité par route (ARA) et y, le pourcentage de la production agricole autoconsommée. D'après le regroupement spatial des résidus des moindres carrés ordinaires (MCO), on limitera notre étude au domaine de corrélation positive $\rho \geq 0$.

A la figure 1, on trace la fonction de vraisemblance pour différentes valeurs de ρ sur l'intervalle (0, 1). Conformément aux résultats établis au chapitre III consacré à l'étude de la fonction de vraisemblance du SAR(1), on retrouve une multitude de maxima locaux autour du maximum global situé en $\hat{\rho} = ,3074$. Chaque segment de la fonction est délimité par un point $\rho = 1/\lambda_i$ où λ_i est la 1ème racine caractéristique de W.

FIGURE 1



Sur quelques sommets, on retrouve la solution finale proposée par l'algorithme de maximisation étant donné divers pas "wt" et différentes valeurs de départ "ρ^{o}" pour ρ. La décomposition de la fonction de vraisemblance aux figures 2 et 3 impute la complexité du problème de maximisation au terme ln ($|P|$), c'est-à-dire au logarithme naturel de la valeur absolue du déterminant de $P = I - \rho W$.

FIGURE 2

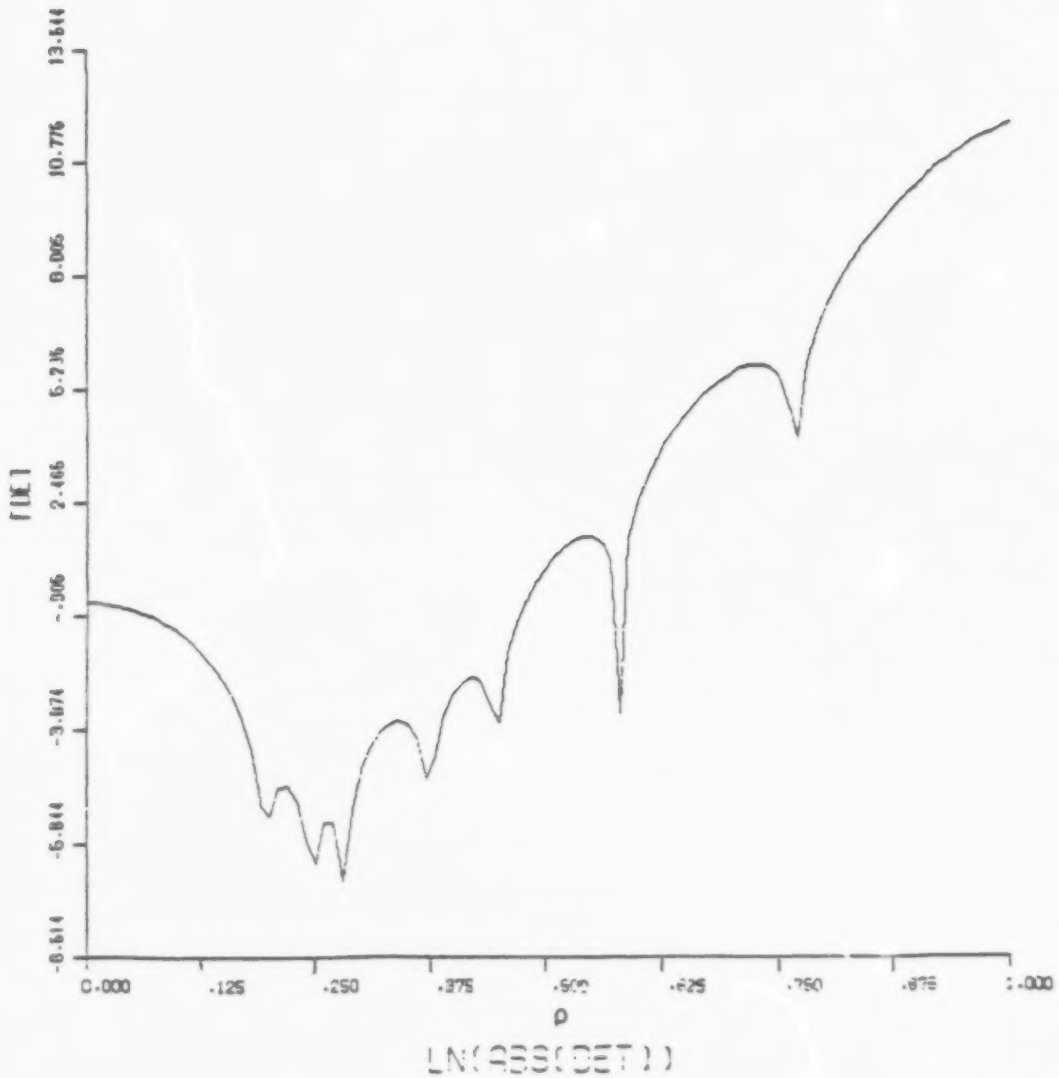
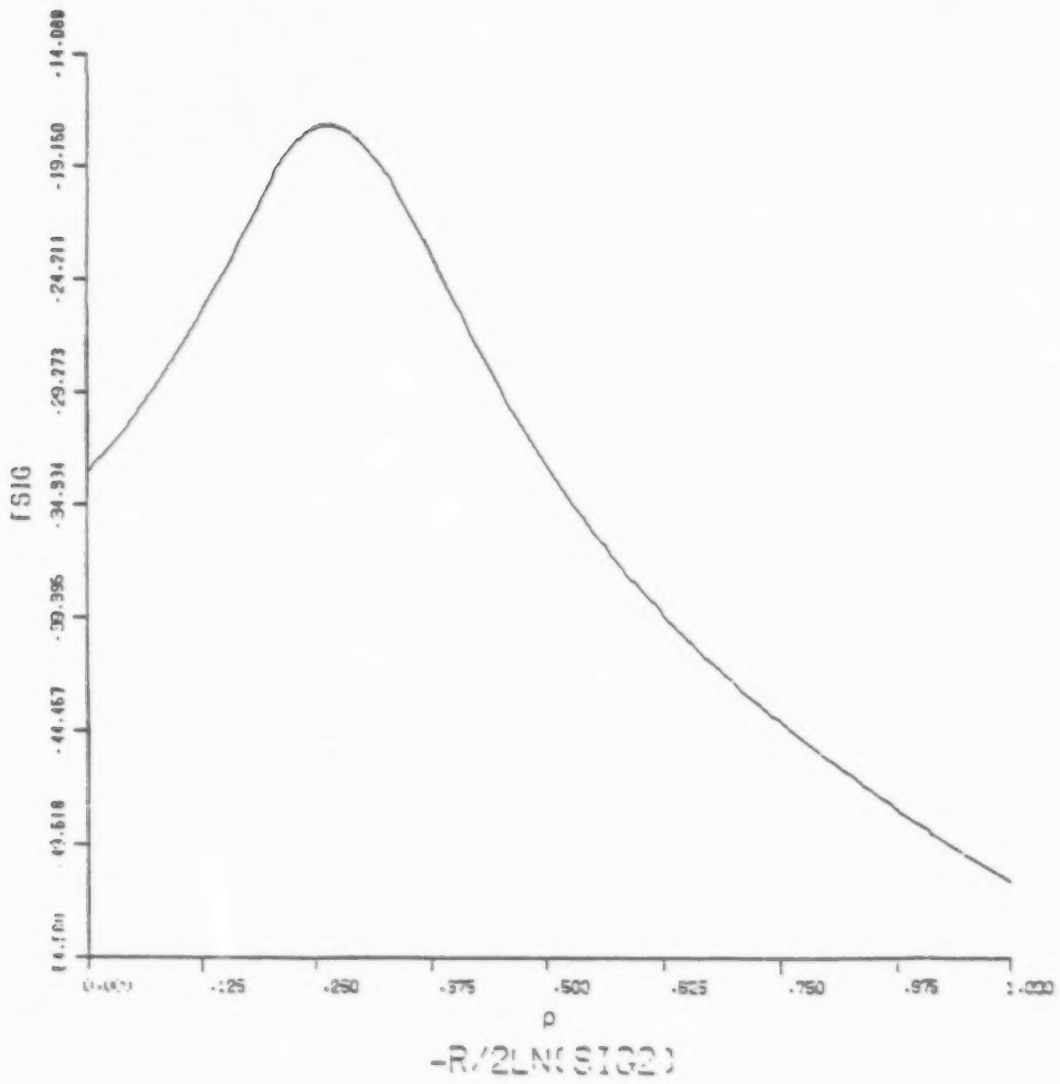

13.844
10.776
8.005
5.236
2.466
-.304
-3.074
-5.844
-8.614
F(K)
0.000
.125
.250
.375
.500
.625
.750
.875
1.000
ρ
LN(ABS(DET))

FIGURE 3


-14.088
-19.150
-24.211
-29.273
-34.334
-39.395
-44.457
-49.518
-54.580
f SIG
.125
.250
.375
.500
.625
.750
.875
1.000
ρ
-R/2LN(SIG2)

A la lumière de la figure 1 précédente, il faut bien se rendre compte que l'algorithme nous a conduit au maximum global qu'une seule fois sur cinq tentatives et qu'il a fallu, pour ce faire, lui fournir une valeur de départ (ρ^0 = ,3) très rapprochée de la valeur optimale. Une telle procédure, sans le traçage de la fonction, semble alors très laborieuse.

D'un point de vue estimation, à la vraisemblance maximale, on obtient les résultats suivants :

$$y = \underset{(0,89075)}{1,97752} + \underset{(6,39190)}{0,00301}\ X$$

(3)

$$\hat{\rho} = \underset{(17,55012)}{0,30736}$$

$R^2 = 0,948856$ $\quad \hat{\sigma}^2 = 4,1697$ $\quad \ln \mathcal{L} = -22,22087$

avec stabilité spatiale si $|\rho| < 0,194897$ [1] .

Les résultats de la régression respectent les attentes, la pente étant positive. Il faut toutefois remarquer que la solution du maximum de vraisemblance se situe en dehors de la région de stabilité spatiale. D'après les résidus $\hat{u}_i$ des moindres carrés généralisés présentés à la figure 4, cette procédure semble avoir modifié de façon marquée la configuration des résidus de la régression linéaire présentée à la figure 3 du chapitre précédent. Bien que l'on ait pu briser certains regroupements de signes, on semble en avoir créé un dans la région sud-est de l'étude.

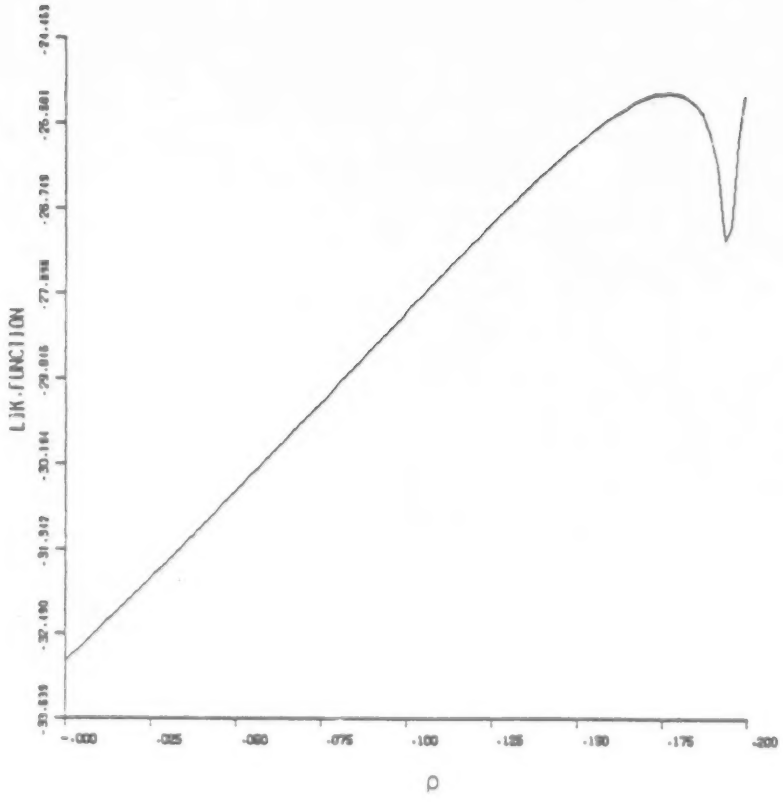
---

1. Comme vu précédemment, pour avoir la stabilité spatiale, il faut respecter : $|\rho| < \frac{1}{|\lambda_{MAX}|}$, où λ_{MAX} est la racine caractéristique maximale de la matrice W.

FIGURE 4

En se limitant à une solution stable, c'est-à-dire en traçant à la figure 5 la fonction sur le domaine [0, 0,195] de ρ, la valeur maximale de $\ln \mathcal{L}(25,2184)$ se situe en $\hat{\rho} = 0,1178$. Quelle que soit la valeur de départ sur cet intervalle, lorsque limité à la solution stable, l'algorithme trouve de façon certaine et ce, en quelques itérations, cette dernière valeur de $\hat{\rho}$.

FIGURE 5



On est alors placé devant un dilemme. En effet, d'un côté, la valeur de $\hat{\rho}$ = ,3074 est, d'après le niveau des vraisemblances, préférable à la solution stable et à la lumière de l'équation (3), conduit à des résultats très satisfaisants. Toutefois, la fonction de vraisemblance sur le domaine [0, 1] possède un comportement très erratique. On peut alors deviner que la tâche deviendra insurmontable lorsqu'on cherchera à estimer, en plus de ρ, les paramètres de la fonction w_{ij}. D'un autre côté, l'unicité de la solution dans l'intervalle de stabilité $|\rho| < 1/|\lambda_{MAX}|$ rend la procédure, contrainte à ce domaine, très attrayante.

---

[1]. On a, en effet, déjà mentionné la forte possibilité d'une seule solution pour $-\frac{1}{|\lambda_{MAX}|} < \rho < \frac{1}{|\lambda_{MAX}|}$.

Pour notre part, comme la stabilité spatiale est un concept qui se veut désirable dans une formulation tirée de la théorie des processus stochastiques et étant donné la simplicité du problème pour trouver une solution dans la région stable de ρ, on adoptera dans les pages qui suivent le principe de la stabilité spatiale[1].

A travers les deux types de normalisation $\underset{\sim}{W}$ et M sur W, on étendra maintenant le domaine de stabilité pour ρ à l'intervalle unitaire, c'est-à-dire que pour $|\rho| < 1$, on trouvera une valeur de ρ qui correspond à un modèle SAR(1) stable. D'après nos précédentes constatations, $\hat{\rho}$ sera fort probablement une solution unique.

2) La forme binaire (0, 1) normalisée

Lorsqu'on parle de matrice de contiguïté normalisée, dans la littérature spatiale, on fait généralement référence à la matrice M formée de la façon suivante :

$$m_{ij} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}}$$

tel que $\sum_{j=1}^{N} m_{ij} = 1$ et où $w_{ii} = m_{ii} = 0 \quad \forall i$. De façon matricielle, soit D une matrice diagonale où l'élément $d_{ii} = \frac{1}{\sum_{j=1}^{N} w_{ij}}$, on a alors :

$$(4) \qquad M = W \cdot D \quad .$$

---

[1]. Comme résultat supplémentaire, si la valeur vraie ρ_0 de ρ appartient au domaine $[-1/\lambda_{MAX}, 1/\lambda_{MAX}]$, lorsque limitée à la solution stable, la procédure du maximum de vraisemblance produit une estimation $\hat{\rho}$ convergente. Voir section 7) du chapitre III.

Même si à l'équation (1), la matrice W est symétrique, la transformation (4) rend la matrice M non symétrique. Comme on veut limiter notre approche à des matrices *symétriques*[1] de contiguïté, on va proposer en 3) une normalisation qui, à la fois respectera la symétrie et maintiendra la condition de stabilité spatiale sur le domaine unitaire $|\rho| < 1$. Mais auparavant, examinons quand même les résultats d'estimation de la régression, dont le schéma autorégressif résiduel est formé par une matrice de contiguïté construite à l'aide de l'équation (4).

2.1) Résultats de l'estimation pour une matrice de contiguïté booléenne normalisée

Bivand (1984) obtient les résultats suivants :

(5) $$y = \underset{(1,51)}{4,671} + \underset{(5,11)}{0,00238}\ X$$

$$\hat{\rho} = \underset{(9,03)}{0,843}$$

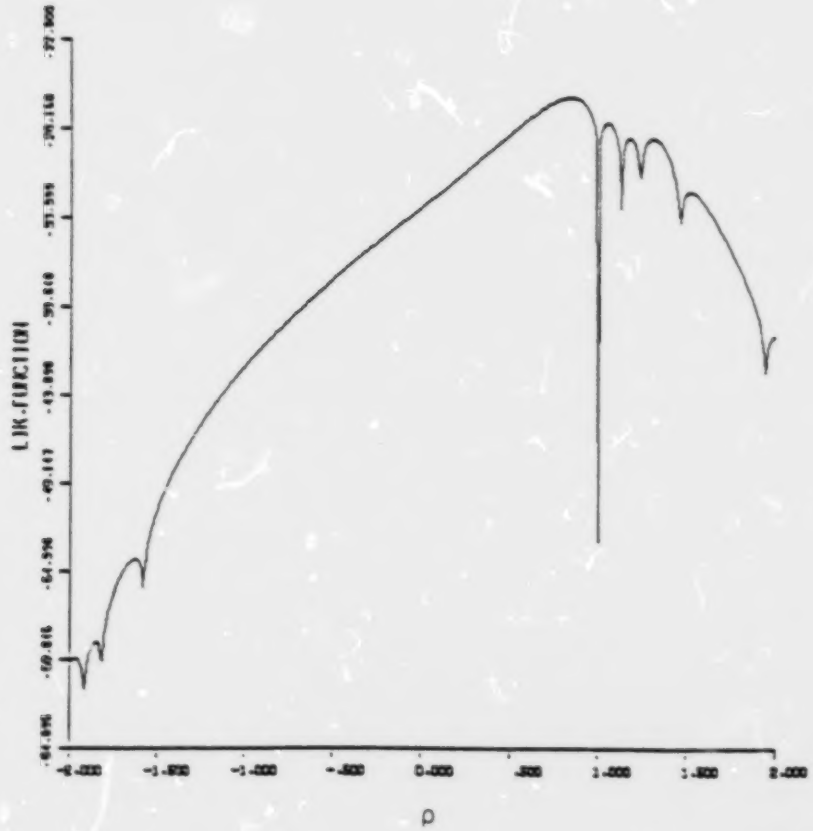
$R^2 = 0,86$ $\qquad \hat{\sigma}^2 = 5,86$ $\qquad \ln \mathcal{L} = -26,40$.

Tout comme pour Burridge (1981), Doreian (1981) ou bien Cliff et Ord (1981), les résultats de cette estimation sont issus d'une procédure de balayage sur la zone permissible (0, 1) de ρ. De la même façon qu'aux estimations précédentes, la constante s'avère non significative. Par contre, l'estimation de la pente est comparable à celles obtenues auparavant et ce de façon significative. La valeur de $\hat{\rho}$ confirme bien la présence d'auto-

[1]. Ainsi que mentionné au début de cet ouvrage, une formulation symétrique, permet du point de vue programmation, une économie substantielle d'espace. C'est un facteur très important dans un problème de type spatial, car, en effet, il faut sauvegarder en mémoire et inverser des matrices de dimension égale à la taille de l'échantillon. Pour un problème qui comporte un grand nombre d'observations, ce facteur devient extrêmement important.

corrélation spatiale résiduelle positive. Après une évaluation sur 200 points sur la région -2 à 2 de ρ, on vérifie bien, à la figure 6, le caractère lisse de la fonction de vraisemblance dans la zone de stabilité. On remarque alors, comme attendu, la présence d'un seul maximum, lequel est imposé par la forme fortement concave de ln $|P|$ entre $-\frac{1}{\lambda_1}$ et 1.

FIGURE 6



L'interprétation économique de la forme (4) est la même que celle de la matrice booléenne standard. En effet, soit le SAR(1) résiduel avec matrice de contiguité M :

(6) $\quad \varepsilon = \rho M \varepsilon + u \ ,$

où, pour chaque observation i, on a :

$$(7) \qquad \varepsilon_i = \rho \sum_{j=1}^{N} m_{ij} \varepsilon_j + u_i \quad .$$

Selon cette formulation, l'erreur de la zone i est fonction d'une somme pondérée des erreurs avoisinantes. Ce poids est en relation inverse avec le nombre de voisins de la zone i, c'est-à-dire que, si une zone i a six voisins, dans la somme qui forme l'erreur ε_i, on accorde à chaque erreur ε_j un poids égal à 1/6.

On l'a dit, l'intérêt principal de cette formulation provient du fait qu'elle permet d'étendre la région de stabilité à l'intervalle $-1 < \rho < 1$. Par contre, la perte de symétrie impliquée par ce type de formulation nous pousse à proposer, à la prochaine section, la normalisation symétrique.

3) La forme binaire (0, 1) avec normalisation symétrique

La normalisation sur la matrice W que l'on va présenter ici permet de conserver à la fois la symétrie et une région permissible intéressante pour ρ. La formulation proposée est la suivante :

$$(8) \qquad \underset{\sim}{w}_{ij} = \frac{w_{ij}}{\sqrt{\sum_{i=1}^{N} w_{ij}} \cdot \sqrt{\sum_{j=1}^{N} w_{ij}}} \qquad \forall i,j \quad .$$

L'unicité probable de la solution et la stabilité dans la zone $-1 < \rho < 1$ sont assurées pour la raison suivante. Soit la normalisation (8) exprimée de façon matricielle :

(9) $\underset{\sim}{W} = D^{1/2} W D^{1/2}$

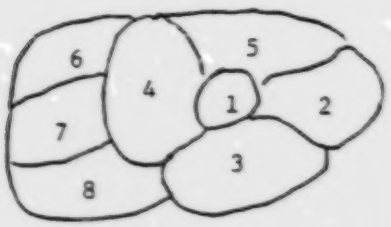
où $d_{ii}^{1/2} = \dfrac{1}{\sqrt{\sum_{j=1}^{N} w_{ij}}}$ et où W est une matrice symétrique.

Comme $\underset{\sim}{W}$ possède les mêmes racines caractéristiques que la matrice normalisée M [1], on conclut que la racine dominante de $\underset{\sim}{W}$ est égale à +1. D'où stabilité lorsque $|\rho| < 1$.

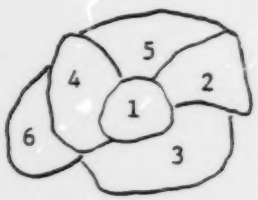
Dans la formulation (7), on a :

$$m_{ij} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}} \qquad \forall i,j \quad .$$

Grâce aux termes $\sqrt{\sum_{i=1}^{N} w_{ij}}$ et $\sqrt{\sum_{j=1}^{N} w_{ij}}$ au dénominateur de (8), on introduit, dans l'étude, de l'information supplémentaire. Pour se convaincre de la pertinence de cette forme, prenons les régions A et B suivantes :



REGION A



REGION B

Intéressons-nous, par exemple, à caractériser l'importance du lien entre les zones 4 et 1. En A, la zone 4 partage son influence avec 6 zones avoisinantes, alors qu'en B, elle ne le fait qu'avec 4 voisins. L'importance de la zone 4 pour la zone 1 doit donc être plus élevée en B.

---

[1]. Les racines caractéristiques d'un produit de matrice ne dépend pas de l'ordre de multiplication des matrices. Donc $\lambda(\underset{\sim}{W}) = \lambda(D^{1/2} W D^{1/2}) = \lambda(WD) = \lambda(M)$.

Dans la formation de l'erreur ε_1, si on utilise l'équation (7), on accordera pour les deux régions le même poids 1/4 à l'erreur ε_4, alors que d'après l'équation (8), tel que désiré, l'erreur ε_4 aura un poids plus élevé dans la région B. En effet, on aura :

$$\text{région A :} \quad \frac{1}{\sqrt{6} \cdot \sqrt{4}} = 0,2041 = w^{A}_{41} \ ,$$

$$\text{région B :} \quad \frac{1}{\sqrt{4}\sqrt{4}} = 0,25 = w^{B}_{41} \ .$$

Donc, en plus de conserver les propriétés intéressantes de symétrie et de stabilité sur le domaine unitaire, cette dernière formulation (8) se révèle, du point de vue de l'interprétation économique, justifiable. On reviendra d'ailleurs sur ce point, au cours des prochaines estimations.

3.1) Résultats de l'estimation pour une matrice de contiguïté booléenne avec normalisation symétrique

L'estimation du modèle de régression qui nous intéresse conduit aux résultats suivants :

(10) $$y = \underset{(1,27515)}{5,28736} + \underset{(4,3991)}{0,00260}\ X$$

$$\hat{\rho} = \underset{(13,62898)}{0,91244}$$

$R^2 = 0,8156$ $\qquad \hat{\sigma}^2 = 4,6449$ $\qquad \ln \mathcal{L} = - 24,41599$.

Ces résultats ressemblent passablement à ceux de la version normalisée (non symétrique) de Bivand (1984) reproduits en 2). L'estimation MCG des paramètres α et β du modèle conduit à des valeurs similaires. En appliquant la normalisation symétrique, la corrélation spatiale positive dans les résidus s'est accentuée. Pour une valeur de départ éloignée du $\hat{\rho}$ optimal, l'algorithme a trouvé la solution en 4 itérations.

D'après la structure des résidus $\hat{u}_i$ des MCG de la figure 7, il semble demeurer, dans les erreurs, des éléments systématiques inexpliqués.

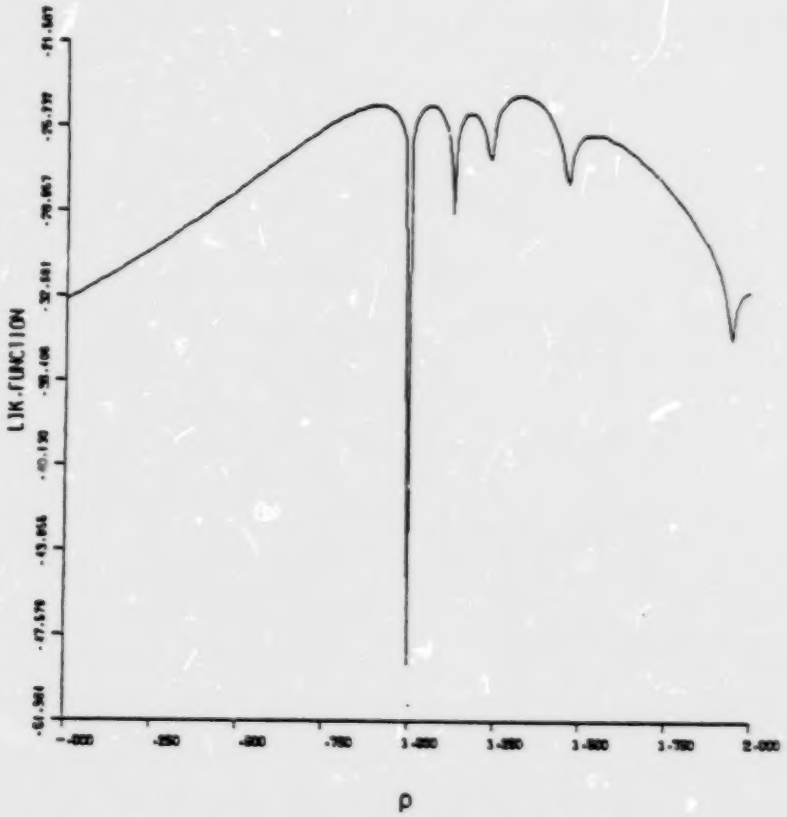
FIGURE 7

Δ changement de signe par rapport à l'estimation MCO

Bien qu'il se soit produit, par rapport à l'estimation MCO, 7 changements de signes, les résidus négatifs semblent encore se regrouper dans la partie sud-est de l'Irlande.

L'évaluation de la fonction de vraisemblance sur 200 points, pour des valeurs de ρ entre 0 et 2 produit la figure 8 suivante :

FIGURE 8



On vérifie l'unicité du maximum sur le domaine stable [0, 1] . On remarque de plus la présence d'un maximum plus élevé ($\ln \mathcal{L} = -23{,}99$) en $\hat{\rho} = 1{,}33$. C'est la solution non stable de la version normalisée symétrique correspondant à la valeur $\hat{\rho} = 0{,}30736$ non stable de la version W standard de la figure 1. On avait souligné à la section 5.3) du chapitre III la présence d'un tel phénomène pour les formulations avec matrice de contiguïté symétrique.

CHAPITRE VII

ESTIMATION DES PARAMÈTRES D'UNE FORME FONCTIONNELLE w_{ij}

Dans les sections qui vont suivre, on tentera d'estimer les paramètres d'une fonction non négative quelconque qui définit w_{ij} :

$$(1) \qquad w_{ij} = g(V^1_{ij}, V^2_{ij}, \ldots, V^P_{ij}, \theta_1, \theta_2, \ldots \theta_H) \quad ,$$

c'est-à-dire une fonction de H paramètres θ_h à estimer et de P variables V^P pertinentes.

Avant de passer à des cas particuliers, essayons d'identifier certains facteurs importants qui puissent expliquer la structure spatiale des résidus.

1) Facteurs qui peuvent expliquer la structure spatiale des résidus

1.1) Autorégression résiduelle

On le rappelle, lorsqu'il y a de l'autocorrélation spatiale résiduelle positive, les erreurs de même signe tendent à se regrouper dans l'espace. Avec de l'autocorrélation négative, une erreur de signe positif (négatif) est généralement entourée d'erreurs de signe négatif (positif). Afin de tenir compte de ce phénomène, Cliff et Ord, avec une formulation autorégressive résiduelle de type SAR(1), proposent d'expliquer l'erreur ε_i par les erreurs environnantes. Pour former cette régression, on peut soit se limiter aux erreurs des zones qui sont physiquement contiguës d'ordre 1 avec la zone i[1], ou bien, comme le suggère le principe de Tobler[2], on peut expliquer l'erreur ε_i par les erreurs de toutes les zones de l'étude.

---

1. Ceci constitue l'approche que l'on a prise au chapitre VI.

2. "The first law of geography : Everything is related to everything else, but near things are more related than distant things" (Cliff et Ord (1981)), p. 8.

Afin d'introduire les caractéristiques spatiales dans le processus autorégressif résiduel, on affecte les erreurs ε_j $(j \neq i)$ qui expliquent ε_i d'un poids quelconque. Pour produire une valeur de ρ qui nous informe sur l'importance et le signe de la corrélation spatiale résiduelle dans l'ensemble de l'étude, ces auteurs proposent, on l'a déjà vu, l'autorégression suivante :

$$\varepsilon_i = \rho \sum_{j=1}^{N} w_{ij}\, \varepsilon_j + u_i \quad .$$

Comme on le voit, le coefficient ρ est commun à toutes les régions. En imposant que les w_{ij} soient non négatifs, avec de l'autocorrélation spatiale résiduelle positive (négative), la valeur de ρ sera > 0 (< 0).

1.2) La distance

Comme l'importance du lien entre deux zones augmente généralement avec leur rapprochement physique, il devient alors pertinent de donner plus de poids aux erreurs des zones les plus rapprochées de la zone ε_i. C'est ce qu'on fera à la section 2) de ce chapitre.

1.3) Autres facteurs

Si, après avoir inclus la distance, il demeure encore de l'autocorrélation spatiale, il faut trouver d'autres facteurs qui puissent justifier cette structure spatiale dans les résidus. Ces facteurs sont généralement propres à l'étude. Cet aspect nous intéressera à la section 3). Pour commencer, prenons le cas le plus simple : la force de contact entre deux zones est une fonction inverse de la distance qui les sépare.

2) $w_{ij} = d_{ij}^{-\theta_1}$

Pour ce cas, notre modèle SAR(1) résiduel s'écrit alors :

(2) $y = \alpha + \beta X + \varepsilon$,

(3) $\varepsilon = \rho W \varepsilon + u$,

où $X = ARA$, $w_{ij} = d_{ij}^{-\theta_1}$ $\forall\, i \neq j$ et $w_{ii} = 0$ $\forall i$ et où d_{ij} représente généralement la distance entre les centres des zones i et j. Selon cette dernière formulation, les éléments non nuls des fonctions w_{ij} ne se limitent pas aux régions contiguës. Ainsi, selon l'équation (3), l'erreur de la zone i est fonction de toutes les erreurs de l'étude.

Une fois la fonction de vraisemblance concentrée sur σ^2, β et α, il reste à maximiser par rapport aux paramètres ρ et θ_1.

En 2.1), on tentera d'estimer les paramètres des équations (2) et (3) pour une matrice de contiguïté W standard. On constatera alors que pour un nombre limité de paramètres dans w_{ij}, il devient extrêmement difficile, même avec des grilles, de trouver les valeurs de ρ et θ_1 qui maximisent la vraisemblance. C'est en 2.2) que l'on s'attardera à la version normalisée symétrique $\underset{\sim}{W}$.

2.1) La forme non normalisée avec $w_{ij} = d_{ij}^{-\theta_1}$

Suite à une grille sur les intervalles suivants : $0 < \rho < 1$ et $0 < \theta_1 < 3$, on a identifié comme sous-région intéressante, la figure 1 qui suit où $0 < \rho < 1$ et $0 < \theta_1 < 1$. On note alors, pour la fonction de vraisemblance, un comportement très erratique. La grille sur la sous-région $,4 < \rho < 1$ et $,2 < \theta_1 < ,5$ nous donne la figure 2.

FIGURE 1

FMAXI, FMINI : -21.43898809383 -32.87319190001

1	.9990	-30.3888	-28.6202	-26.5299	-24.2772	-23.6986	-21.4390	-22.5026	-22.9244	-24.0485	-25.1083
2	.8741	-30.4638	-28.8170	-26.8791	-24.7374	-23.3032	-21.6907	-25.4989	-22.5817	-23.6304	-24.7044
3	.7493	-30.5571	-29.0484	-27.2833	-25.2976	-23.5591	-22.8807	-21.8217	-22.4836	-23.0989	-24.1840
4	.6244	-30.6767	-29.3262	-27.7377	-25.9749	-24.1837	-23.6736	-21.7127	-23.6793	-22.5969	-23.4727
5	.4995	-30.8371	-29.6493	-28.3257	-26.7940	-25.1436	-23.7370	-24.8137	-21.6828	-26.3055	-22.7146
6	.3746	-31.0672	-30.1129	-29.0276	-27.7958	-26.4257	-25.0005	-23.8304	-25.4912	-21.9144	-22.9533
7	.2498	-31.4391	-30.7371	-29.9501	-29.0471	-28.0805	-26.9033	-25.8369	-24.7101	-23.8934	-25.5281
8	.1249	-32.2386	-31.8368	-31.3945	-30.9073	-30.3705	-29.7797	-29.1307	-28.4213	-27.6328	-26.8339
9	0.0000	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732
		0.0000	0556	1111	1667	2222	2778	3333	3889	4444	5000.

1	.9990	-26.4915	-28.4652	-28.4392	-28.3526	-28.2613	-31.8027	-28.9890	-29.1684	-29.9200
2	.8741	-25.8615	-28.1458	-28.0917	-30.2373	-28.2264	-29.0533	-29.4972	-29.0728	-29.5644
3	.7493	-25.2399	-26.6379	-29.1264	-28.6287	-28.8357	-28.5102	-30.3460	-29.2795	-29.3054
4	.6244	-24.5453	-25.6454	-27.2611	-28.7424	-29.3512	-28.8226	-28.8727	-32.2246	-29.3904
5	.4995	-23.5512	-24.6143	-25.7559	-27.3136	-29.2702	-29.4905	-29.2172	-29.0987	-31.7784
6	.3746	-22.8401	-23.1903	-24.1529	-25.2602	-26.5562	-28.7140	-29.3741	-31.8118	-29.3834
7	.2498	-22.2703	-22.0835	-25.1089	-23.1160	-23.7241	-24.7318	-25.9698	-27.5200	-31.2474
8	.1249	-25.9840	-25.1593	-24.4588	-24.1595	-25.7521	-22.7138	-22.1734	-22.7208	-26.0464
9	0.0000	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732
		5556	6111	6667	7222	7778	8333	8889	9444	1.0000

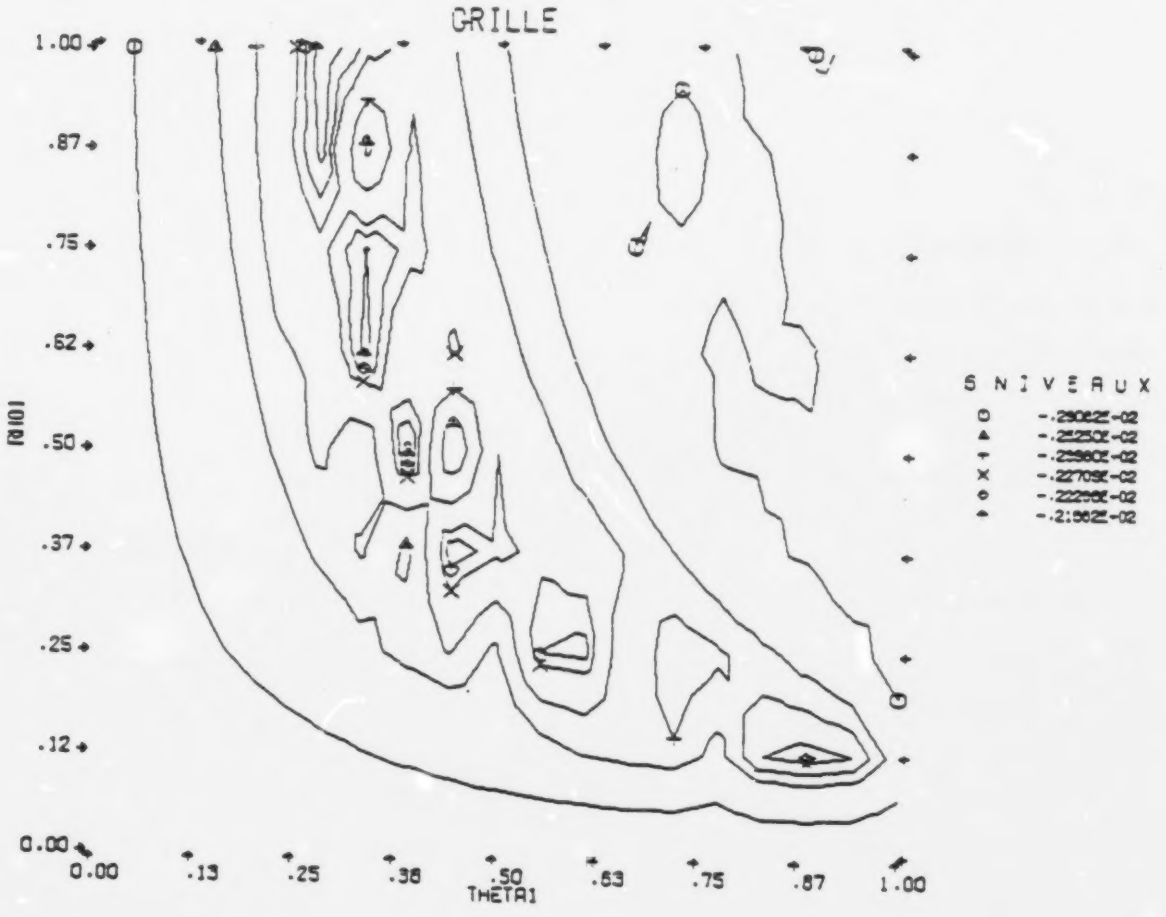
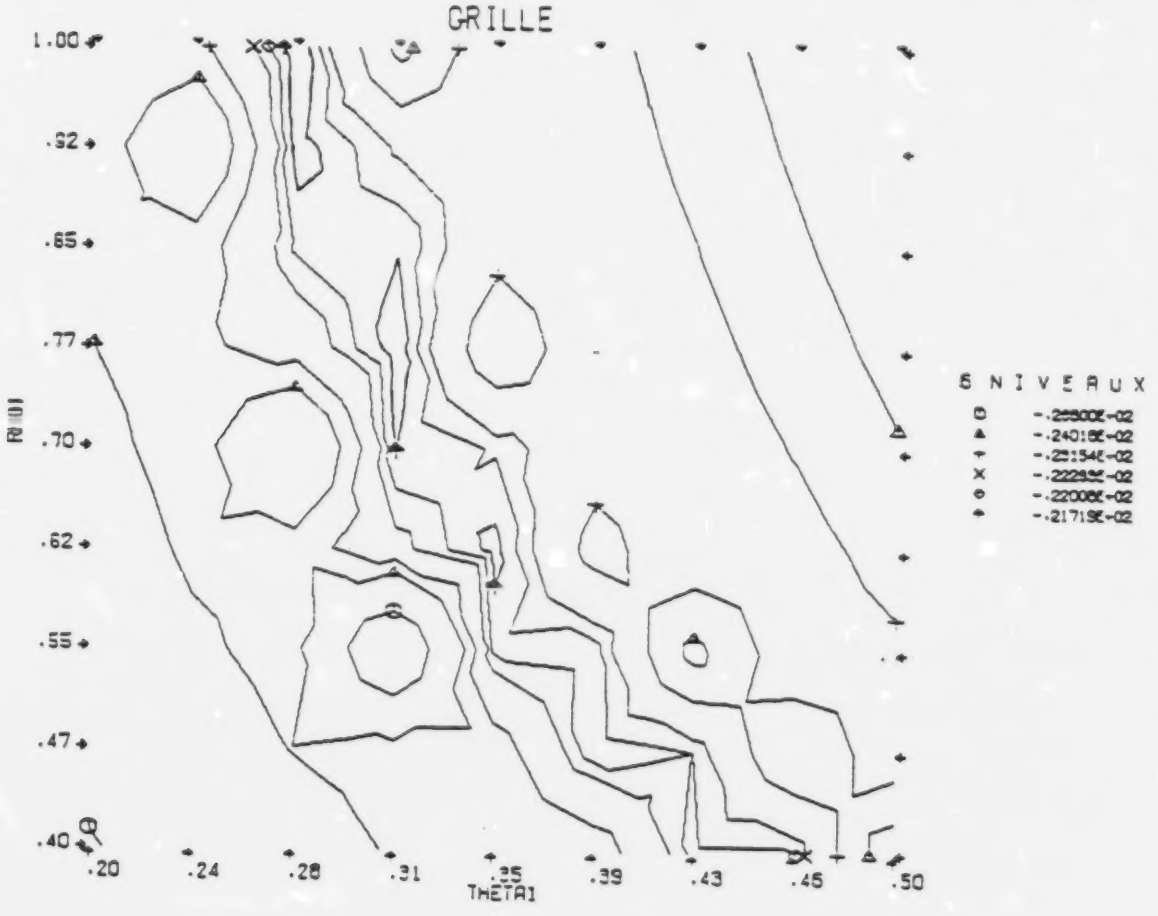


FIGURE 2

FMAXI.FMINI : -21.431623133007 -29.183995318164

1	.9990	-23.3127	-23.3389	-21.4316	-24.2673	-22.3621	-22.8769	-21.6631	-24.3935	-23.1083
2	.9241	-23.4767	-25.3998	-21.5557	-22.3569	-22.4342	-22.6883	-23.4195	-24.1642	-24.8760
3	.8493	-23.7134	-23.3922	-21.9230	-21.7351	-22.9126	-22.5101	-23.1437	-23.8924	-24.6125
4	.7744	-24.0264	-23.3540	-22.7087	-21.5246	-23.8497	-22.4508	-22.8505	-23.3696	-24.3020
5	.6995	-24.4097	-23.4636	-25.9476	-21.7039	-22.0414	-22.8212	-22.5948	-23.1972	-23.9263
6	.6246	-24.8671	-23.7872	-23.5770	-22.4562	-21.6083	-23.4387	-22.5635	-22.8143	-23.4744
7	.5498	-25.4015	-24.2828	-23.3229	-29.1848	-21.9163	-21.8290	-24.2630	-22.5991	-22.9784
8	.4749	-26.0183	-24.9326	-23.9739	-23.5568	-23.6186	-21.8182	-22.0022	-23.9082	-22.6721
9	.4000	-26.7234	-25.7292	-24.7482	-23.9228	-23.6574	-23.6426	-21.9329	-21.9273	-24.6375
		.2000	.2375	.2750	.3125	.3500	.3875	.4250	.4625	.5000



D'après les nombreuses grilles faites au cours des expériences pour une matrice W non normalisée, il s'est avéré généralement très difficile d'identifier de façon certaine une seule portion intéressante pour la recherche d'un optimum global. Le caractère hasardeux de cette procédure de recherche d'un maximum ne sera pas présent lorsqu'on imposera, à la prochaine section, la normalisation symétrique. C'est d'ailleurs ce même phénomène qui explique pourquoi on ne présentera aucun résultat d'estimation dans la présente section.

2.2) $w_{ij} = d_{ij}^{-\theta_1}$ avec normalisation symétrique

2.2.1) *Allure de la fonction de vraisemblance*

Lorsqu'on impose la normalisation symétrique $\underset{\sim}{W}$ à la matrice W de la section précédente, la fonction de vraisemblance revêt alors un comportement beaucoup plus uniforme. A la figure 3, on présente la grille pour les valeurs de ρ et θ_1 sur la région suivante : $-1 < \rho < 1$ et $0 < \theta_1 < 4$. A l'examen de cette figure, il ressort qu'un maximum global stable se situera probablement dans la sous-région $,6 < \rho < 1$ et $2 < \theta_1 < 4$. A la figure 4, on retrouve la grille autour du point maximum situé en $\rho = 0,8005$ et $\theta_1 = 3,3565$. On vérifie alors le caractère très lisse du comportement de la fonction de vraisemblance pour une solution stable.

On l'a vu, ce caractère d'uniformité s'explique de la façon suivante. Pour toute valeur de θ_1 dans $w_{ij} = d_{ij}^{-\theta_1}$, la normalisation $\underset{\sim}{w}_{ij}$ donne une matrice $\underset{\sim}{W}$ qui rend l'évolution de la fonction $\ln |I - \rho \underset{\sim}{W}|$, pour $|\rho| < 1$ très stable.

FIGURE 3

FMAXI,FMINI : -27.40129738442 -39.82012942983

1	.9990	-39.8201	-36.6103	-35.8190	-34.3179	-33.0889	-32.0902	-31.8323	-32.3382	-33.3228
2	.8741	-34.9800	-33.6500	-32.0375	-30.4320	-29.0003	-27.9759	-27.3082	-27.3639	-28.0071
3	.7493	-34.2860	-33.2315	-31.8852	-30.4418	-29.1227	-28.1301	-27.3641	-27.4013	-27.5471
4	.6244	-33.8770	-33.0135	-31.9075	-30.6963	-29.3614	-28.6633	-28.0816	-27.8066	-27.7776
5	.4995	-33.3851	-32.8989	-32.0228	-31.0332	-30.1318	-29.3732	-28.8408	-28.5337	-28.4143
6	.3746	-33.3575	-32.8444	-32.1926	-31.4700	-30.7712	-30.1808	-29.7432	-29.4616	-29.3139
7	.2498	-33.1705	-32.8290	-32.3976	-31.9184	-31.4503	-31.0461	-30.7344	-30.5192	-30.3892
8	.1249	-33.0116	-32.8410	-32.6267	-32.3885	-32.1538	-31.9474	-31.7834	-31.6644	-31.5861
9	0.0000	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732	-32.8732
		0.0000	.5000	1.0000	1.5000	2.0000	2.5000	3.0000	3.5000	4.0000

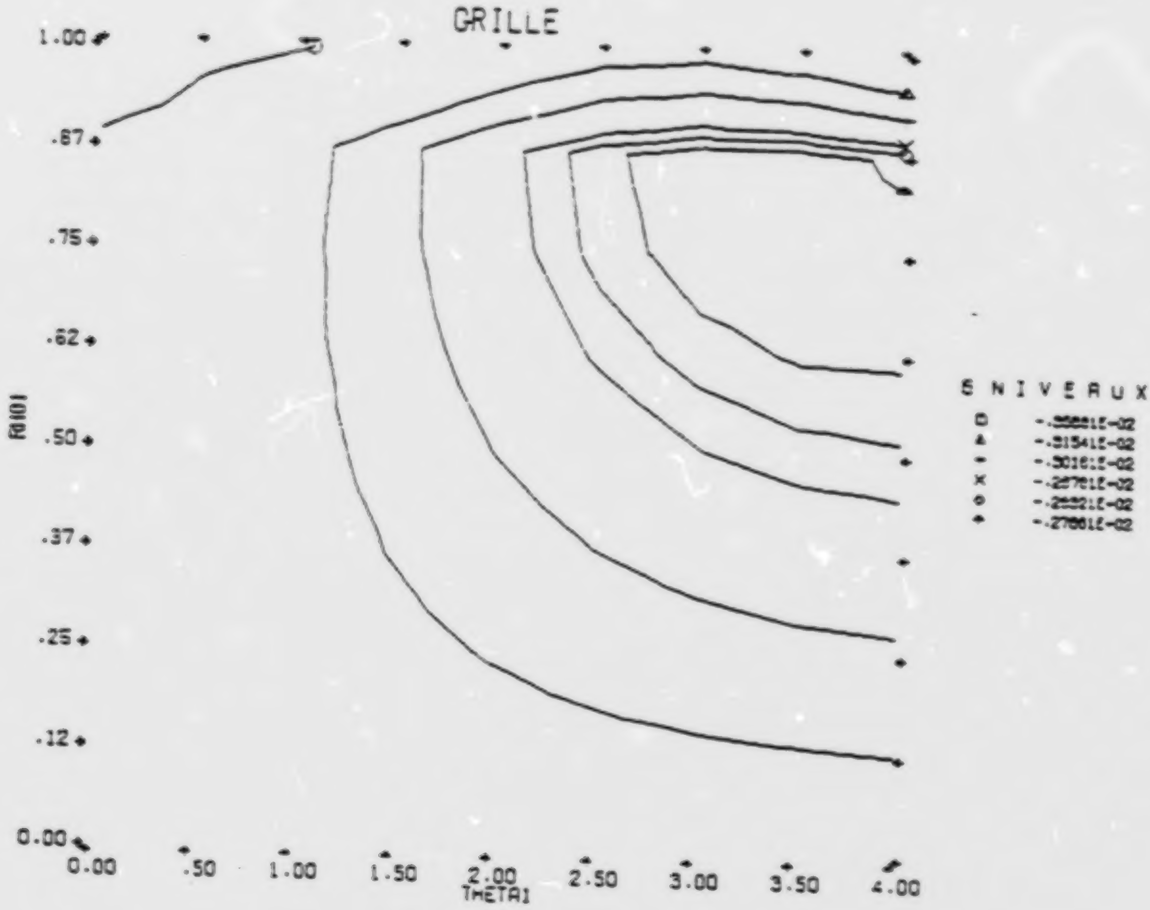
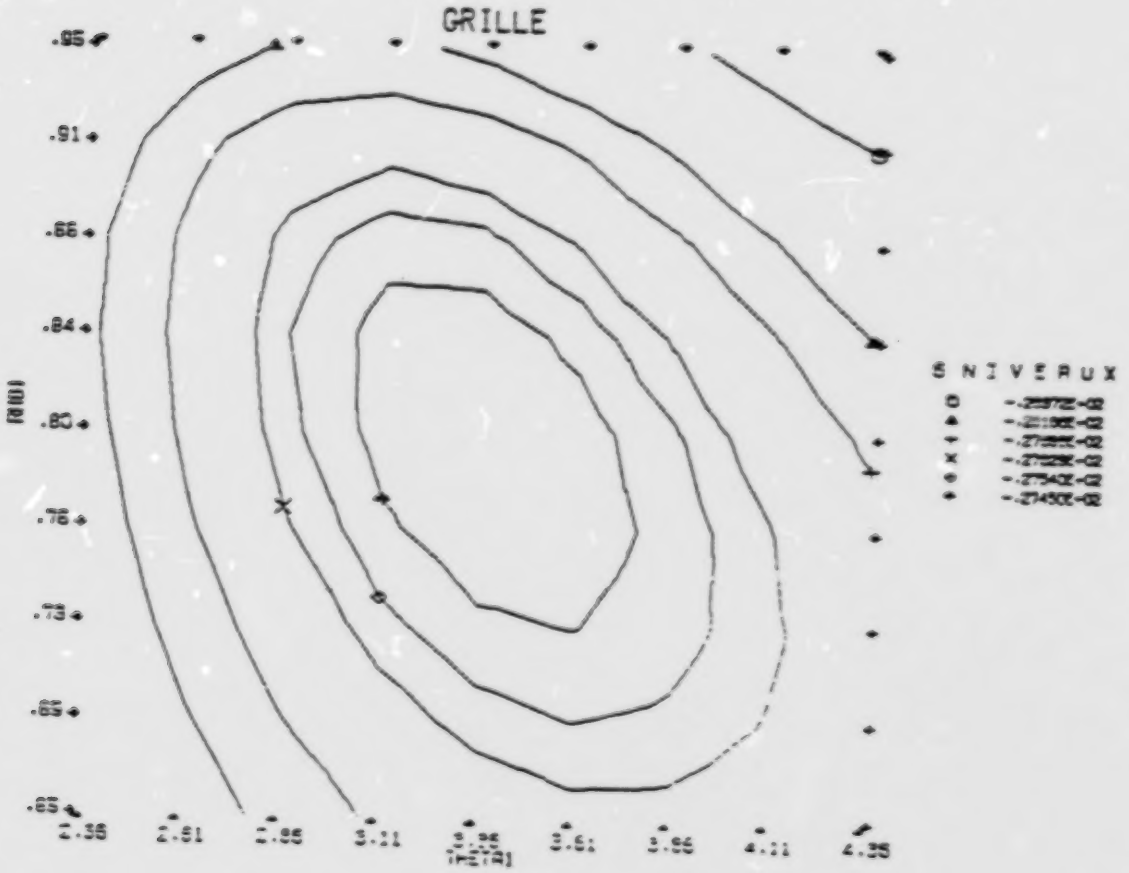


FIGURE 4

FMAX, FMIN : -27.360482360873 -29.777808829134

1	.9505	-28.6790	-28.3146	-28.1341	-28.1039	-28.2260	-28.4954	-28.8952	-29.2899	-29.7778
2	.9130	-28.3415	-27.9625	-27.7434	-27.6773	-27.7496	-27.9389	-28.2216	-28.5752	-29.0003
3	.8755	-28.2182	-27.8304	-27.5873	-27.4847	-27.5089	-27.6411	-27.8605	-28.1475	-28.4849
4	.8380	-28.1990	-27.8040	-27.5422	-27.4095	-27.3928	-27.4783	-27.6548	-27.8737	-28.1552
5	.8005	-28.2436	-27.8442	-27.5674	-27.4101	-27.3607	-27.4048	-27.5261	-27.7090	-27.9293
6	.7630	-28.3222	-27.9215	-27.6440	-27.4661	-27.3887	-27.3984	-27.4811	-27.6225	-27.8098
7	.7255	-28.4322	-28.0346	-27.7399	-27.5663	-27.4436	-27.4463	-27.4956	-27.6018	-27.7509
8	.6880	-28.5993	-28.2058	-27.9076	-27.7022	-27.5829	-27.5393	-27.5603	-27.6346	-27.7316
9	.6505	-28.7652	-28.3798	-28.0813	-27.8682	-27.7344	-27.6709	-27.6680	-27.7134	-27.8038
		2.3563	2.6063	2.8563	3.1063	3.3563	3.6063	3.8563	4.1063	4.3563



Comme le second terme de la fonction de vraisemblance $(-\frac{R}{2} \ln \sigma^{*2})$ [1] a aussi un comportement stable, la combinaison de ces deux facteurs rend la recherche d'un maximum beaucoup plus aisée. C'est ce qui explique pourquoi l'algorithme de maximisation de vraisemblance peut trouver en quelques itérations la solution globale, sur le domaine stable $|\rho| < 1$.

2.2.2) *Interprétation économique*

On a ici :

$$(4) \qquad \underset{\sim}{w}_{ij} = \frac{d_{ij}^{-\theta_1}}{\sqrt{\sum_{i=1}^{N} d_{ij}^{-\theta_1}} \cdot \sqrt{\sum_{j=1}^{N} d_{ij}^{-\theta_1}}} \quad .$$

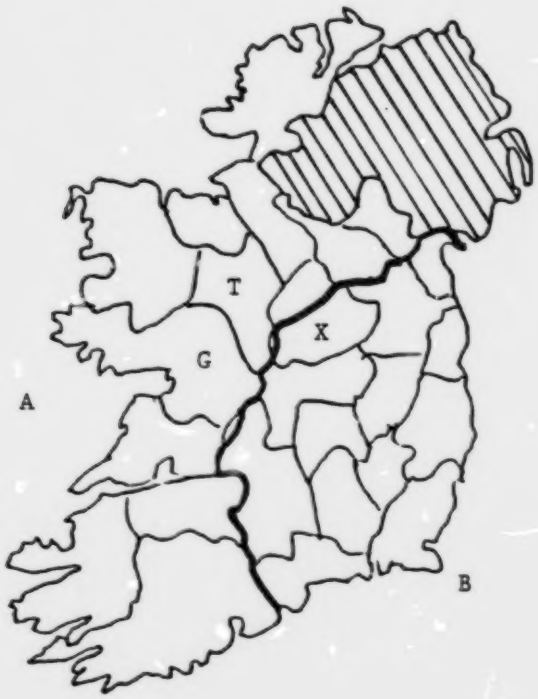
Avec cette forme, le poids $\underset{\sim}{w}_{ij}$ affectant l'erreur ε_j dans la formation de l'erreur ε_i est une relation qui est fonction de la distance entre les zones i et j et des distances de toutes les zones de la région par rapport aux zones i et j concernées.

Pour Cliff et Ord, l'intérêt principal de la matrice W de contiguïté non booléenne provient du fait que l'on peut introduire dans l'étude de l'information sur la structure économico-géographique de la région. Dans le cas précis de l'agriculture en Irlande, O'Sullivan a fait ressortir la présence de deux pôles différents de croissance. Les zones formant la région nord-ouest sont caractérisées par une agriculture de subsistance et, d'après l'indice d'accessibilité (ARA), sont dites éloignées. Au sud-est, on retrouve une concentration de zones limitrophes dont l'activité agricole est tournée vers la commercialisation et l'exportation. Dans le prochain exemple, on va voir que, par rapport aux autres types de normalisation, la forme normalisée symétrique (4) permet de tenir compte de ce phénomène.

---

[1]. Voir chapitre IV, équation (17).

FIGURE 5



	comté	ARA
T	Roscommon	4500
G	Galway	4537
X	Westmeath	3872

distances

	G	X
T	32,81	34,39

D'après la distribution géographique de l'indice d'accessibilité, on peut, à la figure 5, classifier les zones dans l'une des deux catégories suivantes. Dans l'ensemble A, on retrouve les zones éloignées et en B les zones dites rapprochées (faible indice ARA).

Prenons comme référence, la formation de l'erreur ε_T. On remarque que les distances entre les zones T, G et T, X sont relativement semblables. Il se produit alors les deux situations suivantes[1] :

---

[1]. Rappelons que l'indice d'accessibilité d'une zone i est calculé comme suit $\sum_{j=1}^{N} d_{ij}$.

CAS 1 G et T ε A (éloignés)

pour $\theta_1 > 0$, on a donc : $d_{TG}^{-\theta_1}$ élevé $\sum_{i=1}^{N} d_{iG}^{-\theta_1}$ faible

$\sum_{j=1}^{N} d_{Tj}^{-\theta_1}$ faible $\rightarrow w_{TG} = \dfrac{d_{TG}^{-\theta_1}}{\sqrt{\sum_{i}^{N} d_{iG}^{-\theta_1}}\sqrt{\sum_{j}^{N} d_{Tj}^{-\theta_1}}}$ est alors relativement élevé.

CAS 2 X ε B et T ε A

pour $\theta_1 > 0$, on a donc : $d_{TG}^{-\theta_1} \simeq d_{TX}^{-\theta_1}$ élevé $\sum_{i=1}^{N} d_{iX}^{-\theta_1}$ élevé

$\sum_{j=1}^{N} d_{Tj}^{-\theta_1}$ faible $\rightarrow w_{TX}$ a une force mitigée.

D'après nos deux cas, on a donc $w_{TG} > w_{TX}$, ce qui veut dire que dans la formation de l'erreur ε_T, on donnera plus de poids à l'erreur ε_G qu'à l'erreur ε_X, ce qui est désirable si on considère que T et G appartiennent tous deux à la catégorie des zones éloignées. Suite à la remarque d'O'Sullivan sur la présence de deux pôles régionaux de croissance agricole, la formulation (4) se révèle très judicieuse. C'est un avantage face aux deux autres types de normalisation[1] qui accorderaient aux erreurs ε_X et ε_G relativement la même pondération.

---

[1]. On parle évidemment de la forme non normalisée $w_{ij} = d_{ij}^{-\theta_1}$ et de la forme normalisée standard $m_{ij} = \dfrac{d_{ij}^{-\theta_1}}{\sum_{j=1}^{N} d_{ij}^{-\theta_1}}$.

2.2.3] *Résultats de l'estimation*[1]

$$y = \underset{(-0,68637)}{-3,03516} + \underset{(5,6978)}{0,00414}\ X \quad ,$$

$$\hat{\rho} = \underset{(5,21509)}{0,8004679} \quad , \tag{5}$$

$$\hat{\theta}_1 = \underset{(3,73787)}{3,356529} \quad ,$$

$R^2 = 0,8014 \qquad \hat{\sigma}^2 = 6,71197 \qquad \ln \mathcal{L} = -27,36068$.

La valeur de $\hat{\rho}$ confirme de façon significative la présence d'autocorrélation positive forte dans les résidus du modèle MCO. D'après l'évidence statistique du test t sur θ_1, il semble que le degré de lien entre les erreurs est une fonction non linéaire de la distance, où on a $d_{ij}^{-3,36}$.

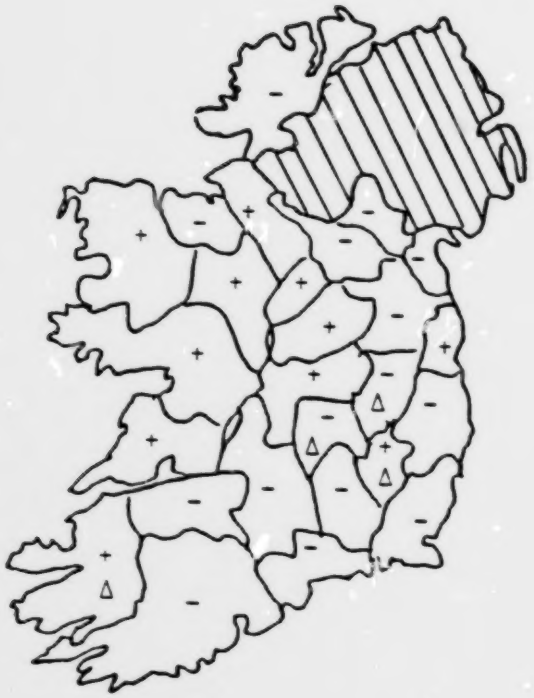
La pente de la régression a une valeur tout aussi raisonnable que lors des estimations précédentes et, de la même façon qu'auparavant, la constante s'avère non significative. L'algorithme a trouvé cette solution en neuf itérations et par les grilles de la section 2.2.1), on a pu se convaincre de l'unicité de cette solution sur le domaine $|\rho| < 1$.

A la figure 6, on reproduit la configuration spatiale des résidus $\hat{u}_i$ observés. Cette estimation semble avoir été incapable de rendre aléatoire les résidus du modèle transformé. Les signes des erreurs ont conservé relativement les mêmes emplacements.

---

[1]. On ne présentera pas les résultats d'estimation pour les formes normalisées de $w_{ij} = (\ln d_{ij})^{-\theta_1}$ et $w_{ij} = e^{-\theta_1 d_{ij}}$ car elles ont conduit à des résultats relativement similaires à ceux présentés dans cette section.

FIGURE 6



Δ changement de signe par rapport à l'estimation MCO

D'après le niveau de la fonction de vraisemblance, il semble plus raisonnable, dans la formation du processus autorégressif (3), de se limiter qu'aux erreurs de zones contiguës. En effet, dans les estimations précédentes où θ_1 était fixé à zéro, l'erreur ε_i n'était fonction que des erreurs ε_j contiguës d'ordre 1. On obtenait alors une valeur plus élevée de la fonction de vraisemblance. C'est ce que l'on va tester à la section 2.3) suivante.

2.3) $w_{ij} \neq 0$ si i et j sont contiguës

Dans le cas $w_{ij} = d_{ij}^{-\theta_1}$, avant d'exécuter une quelconque sorte de normalisation, les deux types de matrices de contiguïté d'ordre 1 qui retiennent généralement l'attention sont :

$$(6) \qquad w_{ij} \begin{cases} = d_{ij}^{-\theta_1} & , \quad \text{pour tout } i \neq j \ , \\ = 0 & , \quad \text{si } i = j \ , \end{cases}$$

et

$$(7) \qquad w_{ij} \begin{cases} = d_{ij}^{-\theta_1} & , \quad \text{si i est contiguë à j d'ordre 1} \ , \\ = 0 & , \quad \text{sinon} \ . \end{cases}$$

Définissons une matrice G où :

$$(8) \qquad g_{ij} \begin{cases} = 1 & , \quad \text{si i et j sont contiguës} \ , \\ = 0 & , \quad \text{ailleurs} \ . \end{cases}$$

On peut alors inclure les formulations (6) et (7) dans la forme suivante :

$$(9) \qquad w_{ij} = d_{ij}^{-\theta_1} (\theta_3 + \theta_2 g_{ij}) \qquad \theta_3 \geq 0 \qquad \theta_2 \geq 0 \ [1].$$

Pour $\theta_2 = 0$ et $\theta_3 = 1$, on a l'équation (6) et pour $\theta_2 = 1$ et $\theta_3 = 0$, on a l'équation (7).

---

[1]. Afin de laisser à ρ son interprétation sur le signe de l'autocorrélation spatiale résiduelle, les w_{ij} doivent être des poids non négatifs. C'est ce qui nous oblige à contraindre les paramètres θ_2 et θ_3.

Dans un modèle plus général, on tentera d'estimer la valeur des paramètres θ_1, θ_2 et θ_3. Dû à un problème d'identification entre les paramètres θ_2 et θ_3, lors d'une normalisation symétrique, le modèle général que l'on estimera s'écrit[1] :

$$w_{ij} = d_{ij}^{-\theta_1} (\theta_3 + g_{ij}) \quad , \tag{10}$$

où encore, de façon équivalente :

$$w_{ij} = d_{ij}^{-\theta_1} (1 + \theta_2 g_{ij}) \quad . \tag{10.1}$$

Pour $\theta_2 = 0$, l'estimation à la section 2.2.3) précédente nous a donné $\ln \mathcal{L}_{\omega_1} = -27{,}36068$. Pour $\theta_3 = 0$, on obtient, on le verra, $\ln \mathcal{L}_{\omega_2} = -24{,}382085$. Pour sa part, l'estimation jointe de θ_1 et θ_3 à l'équation (10) produit une valeur $\hat{\theta}_3 = 0$, ce qui produit une valeur $\ln \mathcal{L}_\Omega = \ln \mathcal{L}_{\omega_2}$. L'algorithme atteint en fait une solution de coin car les valeurs négatives de θ_3 lui sont interdites. Comme pour passer de l'estimation du modèle (6) [$\theta_2 = 0$] au modèle (10.1), il faut relâcher une contrainte ($\theta_2 = 0$), le test du rapport des vraisemblances se fera avec un degré de liberté.

---

1. Pour $w_{ij} = \gamma\, d_{ij}^{-\theta_1}$, on a :

$$\underset{\sim}{w}_{ij} = \frac{\gamma}{\gamma} \frac{d_{ij}^{-\theta_1}}{\sqrt{\sum_i d_{ij}^{-\theta_1}} \sqrt{\sum_j d_{ij}^{-\theta_1}}}$$

ce qui revient à dire que toute valeur de γ est correcte. Fixons alors γ à 1. Comme l'équation (9) peut s'écrire :

$$w_{ij} = \theta_2\, d_{ij}^{-\theta_1} \left(\frac{\theta_3}{\theta_2} + g_{ij}\right) ,$$

le fait de fixer $\theta_2 = 1$ conduit directement à l'équation (10). Pour $\theta_3 = 1$, on a l'équation (10.1).

Comme $-2\ [\ln \mathcal{L}_{\omega_1} - \ln \mathcal{L}_{\Omega}] \sim \chi^2_{(1)}$, on a alors[1] :

$$2\ [-\ 24{,}382085 + 27{,}36068] = 5{,}957719 > 3{,}841 \quad .$$

Le modèle dont la formulation autorégressive ne dépend que des résidus contigus semble, d'après l'évidence statistique de ce test, plus vraisemblable.

2.3.1) *Estimation du modèle* $w_{ij} = d_{ij}^{-\theta_1}$ *, si i et j sont contigus avec normalisation symétrique*

L'estimation de ce modèle conduit aux résultats suivants :

$$y = \underset{(1{,}1773)}{5{,}04134} + \underset{(4{,}40716)}{0{,}00266}\ X \quad ,$$

(11) $$\hat{\rho} = \underset{(13{,}2987)}{0{,}909322} \quad ,$$

$$\hat{\theta}_1 = \underset{(0{,}2954121)}{0{,}281913} \quad ,$$

$R^2 = 0{,}817626$ $\qquad$ $\sigma^2 = 4{,}64675$ $\qquad$ $\ln \mathcal{L} = -\ 24{,}382085$.

Dans une formulation où les erreurs ne dépendent que des erreurs des zones contiguës, la distance perd sa signification. En effet θ_1 n'est pas significativement différent de zéro. C'est ce qui explique la similitude de ces résultats avec ceux présentés à la section 3.1) du chapitre précédent.

---

[1]. Soit $\lambda = \mathcal{L}_{\omega_1} / \mathcal{L}_{\Omega}$ où ω_1 représente le modèle contraint, $-\ 2 \ln \lambda \sim \chi^2_{\nu}$ avec ν restrictions. Le test du rapport de vraisemblance est un test bilatéral. D'après l'équation (9) où θ_2, $\theta_3 \geq 0$, un test unilatéral serait plus puissant que ce dernier. La valeur calculée du t de θ_3 à l'équation (10) étant beaucoup trop élevée, on préfère ici le test moins puissant. Le problème n'est toutefois pas sérieux car l'évidence empirique du test (5,96 > 3,84) nous pousse tout de même à rejeter H_o.

Dans le cas de la régression qui nous concerne ici, quelle que soit la forme fonctionnelle w_{ij} sélectionnée, lorsqu'on décide d'inclure dans la forme autorégressive seulement les erreurs contiguës, seul le ρ s'avère significatif. A la lumière des estimations présentées jusqu'ici, lorsque dans le processus régressif résiduel, on veut ne tenir compte que des erreurs de zones contiguës, il est préférable de choisir, comme matrice de contiguïté, la forme booléenne (0, 1) normalisée symétrique.

3) Forme généralisée $w_{ij} = d_{ij}^{-\theta_1} \cdot p_{ij}^{\theta_2}$ [1]

A la section 2.3) précédente, où $w_{ij} = d_{ij}^{-\theta_1}$ $\forall i \neq j$, on a obtenu un θ_1 significatif. Toutefois, par l'analyse des résidus de la figure 6 correspondante, on voit qu'il demeure sûrement encore de l'autocorrélation résiduelle positive[2]. Afin de briser cette configuration résiduelle, il faut trouver des variables qui vont expliquer ce regroupement. Pour ce faire, réécrivons le processus autorégressif (3) comme suit :

$$\varepsilon = \rho \tilde{\varepsilon} + u \qquad (12)$$

Cette équation représente donc une régression à une seule variable explicative $\tilde{\varepsilon}$ formée à partir d'une somme quelconque des erreurs, où chacun des termes ε_j est pondéré par un certain poids. Ces dernières pondérations sont habituelle-

---

[1]. Afin d'éviter des contraintes de non négativité sur les paramètres d'une forme linéaire de type $d_{ij}^{-\theta_1} (\theta_2 + \theta_3 p_{ij}^{\theta_4} + \ldots.)$, on se limitera pour la forme généralisée à une spécification similaire à celle employée dans les modèles de gravité, c'est-à-dire une formulation multiplicative.

[2]. En l'absence d'un test formel de validation "diagnostic checking test", une méthode visuelle est ici employée.

ment sélectionnées afin de décrire la structure spatiale propre de la région. Jusqu'ici, on a donné les poids les plus grands aux erreurs les plus rapprochées[1].

Le caractère décevant de ce dernier exercice nous suggère d'ajouter à la distance des variables qui peuvent servir à décrire la structure spatiale des résidus MCO.

FIGURE 7

FIGURE 8

[1]. Cette équation (12) prend la même signification dans le domaine des séries temporelles. En effet, pour un AR(1) résiduel, on décide de donner un poids égal à 1 à l'observation précédente (t-1) et un poids égal à zéro à toutes les autres. Dans les applications économiques, cette spécification suffit généralement à expliquer les éléments systématiques présents dans le vecteur d'erreur ε. Si cette démarche s'avère insuffisante, la procédure en 3 étapes établie par Box et Jenkins suggère de respécifier l'ordre d'autorégression dans les résidus jusqu'à ce que les u_i soient suffisamment aléatoires.

On peut remarquer à la figure 8 que le comportement du logarithme naturel du revenu ressemble à celui des résidus MCO (Figure 7). On a donc cherché à ajouter cette variable dans la spécification du modèle $y = \alpha + \beta X + \varepsilon$. Cette tentative n'ayant pas produit les résultats escomptés on a essayé la variable binaire $z = 1$ lorsque $\ln (Rev) > 9,5$ et $z = 0$ ailleurs. N'ayant pas produit un meilleur résultat, on tente maintenant d'introduire ce facteur dans la structure des résidus. D'après les figures 7 et 8 précédentes, dans la formation de l'erreur ε_i, on aimerait donner plus de poids aux erreurs ε_j dont le log du revenu $\ln Rev_j$ ressemble au $\ln Rev_i$. En examinant la série, on s'aperçoit que de façon générale, lorsque les zones i et j ont toutes deux des petits revenus $|\ln Rev_i - \ln Rev_j|$ est faible, lorsque les zones i et j ont de grands revenus, $|\ln Rev_i - \ln Rev_j|$ est faible et lorsqu'un des revenus est grand et que l'autre est petit $|\ln Rev_i - \ln Rev_j|$ est fort. Ce fait nous incite à proposer comme facteur supplémentaire à la distance la variable suivante :

$$(13) \qquad p_{ij} = \frac{1}{|\ln Rev_i - \ln Rev_j|}$$

C'est l'inverse numérique de la valeur absolue de la différence des logarithmes naturels du revenu pour les zones i et j concernées.

3.1) Interprétation économique

On peut remarquer, d'après la figure 8, que les zones dont les erreurs sont positives sont généralement celles qui possèdent les plus petits revenus et de la même façon les zones négatives ont de forts revenus. Il semble alors

pertinent, dans le schéma autorégressif, de donner plus de poids aux erreurs qui appartiennent à la même classe de revenu. C'est ce que permet la formulation (13). En effet, *de façon générale*, d'après le paragraphe précédent, si la zone i et la zone j ont toutes deux un petit revenu, p_{ij} est alors élevé. Si Rev_i est grand alors que Rev_j est petit (ou l'inverse), p_{ij} devient petit. Pour des Rev_i et Rev_j grands, p_{ij} est grand.

3.2) Résultats d'estimation pour $w_{ij} = d_{ij}^{-\theta_1} \cdot \left(\frac{1}{|\ln Rev_i - \ln Rev_j|}\right)^{\theta_2}$

En fait, ce que l'on estime, c'est la version normalisée symétrique : $w_{ij} \neq 0$ si $i \neq j$, on obtient alors, en 10 itérations :

$$y = \underset{(-0,0629)}{-0,24215} + \underset{(6,05842)}{0,00394}\ X \quad ,$$

$$\hat{\rho} = \underset{(8,888314)}{0,0867935} \quad , \tag{14}$$

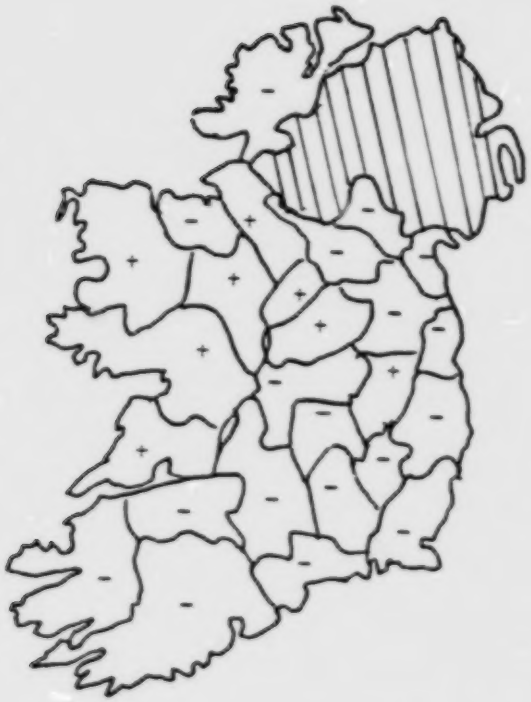
$$\hat{\theta}_1 = \underset{(3,497420)}{4,004878} \quad ,$$

$$\hat{\theta}_2 = \underset{(2,24435)}{0,532371} \quad ,$$

$R^2 = 0,87149$ $\qquad \hat{\sigma}^2 = 5,77144$ $\qquad \ln \mathcal{L} = -26,590896$.

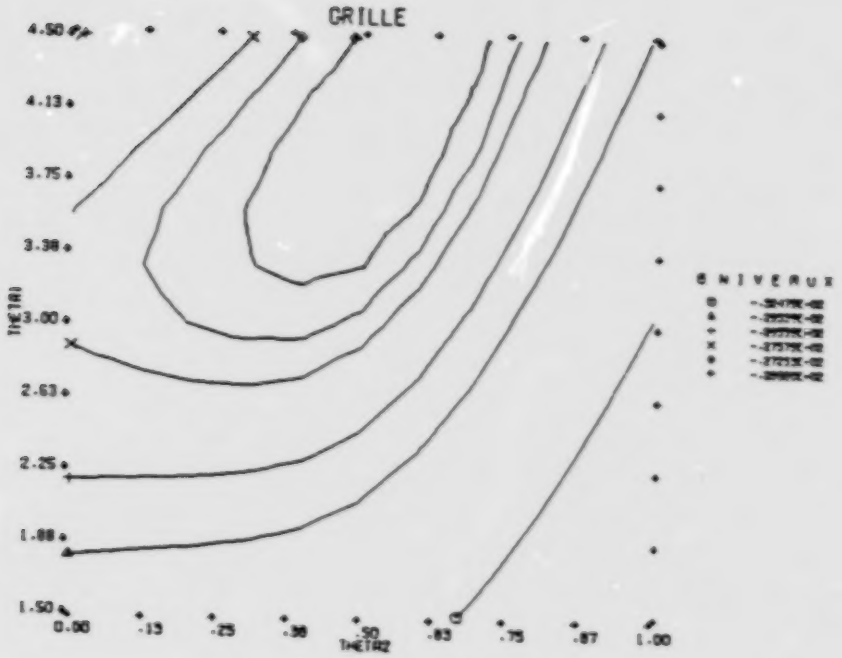
Comparativement à l'estimation de la section 2.2.3) précédente, seule la constante a bougé. La pente donne une valeur tout aussi raisonnable qu'auparavant. Ce qui est le plus pertinent, dans la dernière estimation, c'est le niveau de signification de ρ, θ_1 et θ_2. En effet, ces trois paramètres, en plus d'être du bon signe, sont significatifs. La fonction de vraisemblance n'a toutefois pas assez diminué pour permettre de conclure en la pertinence d'introduire la variable (13).

FIGURE 9



D'après la figure 9, cette estimation semble avoir détérioré la situation. Comme, dans la zone stable, le $\hat{\rho}$ est unique, traçons à l'aide d'une grille le comportement de la fonction de vraisemblance pour différentes valeurs de θ_1 et θ_2. Pour ce faire, fixons $\rho = 0{,}867935$.

FIGURE 10



Le caractère lisse qu'implique la normalisation symétrique sur $\ln |I - \rho \underset{\sim}{W}|$ explique le caractère uniforme des isocourbes. Ajoutons de plus que l'algorithme a trouvé en 10 itérations la solution présentée à l'équation (14). L'estimation de cette forme w_{ij} pour le cas contigu produit des valeurs de θ_1 et θ_2 non significatives.

Les autres variables que l'on a créées se sont toujours révélées non significatives. On doit donc avouer que l'on n'a pu, grâce à l'estimation des paramètres de la forme w_{ij}, briser la corrélation présente dans les résidus. Ceci n'est pas surprenant si on considère que les procédures alternatives employées par Bivand, Cliff et Ord, Burridge, Blommestein ont donné les mêmes résultats.

Pour briser cette configuration résiduelle, il faudrait peut-être passer à un ordre supérieur d'autorégression. Rappelons toutefois que l'objectif principal de cette thèse était de développer une méthode d'estimation pour les paramètres de la matrice de contiguïté d'un SAR(1) résiduel. Il sera sûrement très intéressant de reprendre l'exercice sur d'autres problèmes que l'on espère un peu moins chroniques.

On aimerait finalement rappeler que parmi les expériences tentées, seule la solution non stable avec matrice de contiguïté booléenne non normalisée a semblé briser de façon plus satisfaisante la structure autocorrelée des résidus. Les résultats de la section 7) au chapitre III suggèrent toutefois que la procédure du maximum de vraisemblance produit dans ce cas des estimations non convergentes.

CHAPITRE VIII

FORMULATION AUTORÉGRESSIVE POUR DES DONNÉES DE FLUX

Soit la régression suivante :

$$(1) \qquad Y_{ij} = \beta_0 + X_{ij}^1 \beta_1 + \ldots + X_{ij}^k \beta_k + \xi_{ij} \quad , \qquad i = 1,\ldots,M;\ j = 1,\ldots,N \ ,$$

où on veut expliquer des flux Y entre les paires de zones i et j. Selon cette formulation, on possède alors M zones en origine et N zones de destination. Tout dépendant du problème, l'ensemble des zones à l'origine I peut être soit identique, soit distinct de l'ensemble J que forme la destination. En effet, dans le premier cas, l'équation (1) peut servir à expliquer les migrations entre les zones de deux régions avoisinantes. Selon la seconde formulation, "I ≡ J", ce sont les déplacements entre les zones d'une seule région qui nous intéressent.

Formulation du schéma autorégressif

Afin de caractériser de façon satisfaisante les liaisons entre les erreurs ξ_{ij}, Brandsma et Ketellapper (1979a) proposent d'utiliser le principe de contiguïté si populaire dans la recherche spatiale. En effet, pour conserver l'idée de contiguïté, définissons une matrice C de contiguïté en origine et une matrice B en destination. Soit ξ_{rs} l'erreur de régression pour le flux entre la zone r et la zone s. On a alors M zones adjacentes en origine qui compétitionnent avec r pour aller vers s, et N zones en destination qui rivalisent avec s pour attirer ce qui vient de r. Posons :

$$\underset{(N\times N)}{B} = B' \quad , \quad \text{matrice symétrique de contiguïté en destination,}$$

$$\underset{(M\times M)}{C} = C' \quad , \quad \text{matrice symétrique de contiguïté en origine,} \qquad (2)$$

$$\underset{(M\times N)}{\xi} = [\xi_{rs}] \quad , \quad \text{matrice M×N des erreurs de l'équation (1)}$$

avec $\xi_{i.}$ la ligne i de ξ et $\xi_{.j}$ la colonne j de ξ .

Le processus autorégressif proposé par Brandsma et Ketellapper (1979a) s'écrit alors :

$$\xi_{rs} = \rho_1 \xi_{r.} B_{.s} + \rho_2 C_{r.} \xi_{.s} + \zeta_{rs} = \rho_1 \sum_{j=1}^{N} \xi_{rj} B_{js} + \rho_2 \sum_{i=1}^{M} C_{ri} \xi_{is} + \zeta_r \qquad (3)$$

$$r = 1,\ldots,M; \quad s = 1,\ldots,N \ ,$$

où $B_{ss} = C_{rr} = 0 \quad \forall r,s$.

Comme les erreurs ξ_{rs} de l'équation sont des flux inexpliqués, dans la justification théorique et économique de la formulation (3), on parlera de flux. Ainsi, d'après (3), l'erreur ξ_{rs} est formée par une somme de tout ce qui part de r vers une destination j autre que s (ξ_{rj}) dont le poids B_{js} exprime le phénomène de compétition en destination, plus une somme pondérée de tout ce qui se dirige vers la destination s en provenance d'une origine i différente de r (C_{ri} représente alors la compétition en origine) plus un terme d'erreur aléatoire ζ_{rs}. On définit par la suite ρ_1 comme le coefficient d'autocorrélation spatiale en destination et ρ_2 celui de l'origine.

Si on répète l'équation (3) pour chaque paire (r,s), on obtient en écriture matricielle :

$$\underset{(M\times N)}{\xi} = \rho_1 \underset{(M\times N)}{\xi} \cdot \underset{(N\times N)}{B} + \rho_2 \underset{(M\times M)}{C} \cdot \underset{(M\times N)}{\xi} + \underset{(M\times N)}{\zeta} \quad . \tag{4}$$

Les éléments des matrices B et C sont, comme d'habitude, des fonctions fixées à priori qui permettent d'introduire les caractéristiques spatio-économiques propres à l'étude. Comme dans les chapitres précédents, on propose d'estimer les paramètres de ces matrices.

L'équation (3) semble à première vue justifiable du point de vue économique. Dans des travaux subséquents, on pourra approfondir les recherches sur la pertinence économique d'une telle formulation. En se limitant ici à l'intérêt de la forme mathématique de (3) et (4), si on vectorise l'équation (4), on obtient un schéma régressif semblable à un SAR(2) résiduel.

En effet, soit :

$$\text{vec}\,(ABC) = (A \otimes C')\,\text{vec}\,(B) \tag{5}$$

Pour B et C symétrique à l'équation (4), on obtient :

$$\underset{(NM\times 1)}{\text{vec}\,(\xi)} = \rho_1 \underset{(NM\times 1)}{\text{vec}\,(I_M \xi B)} + \rho_2 \underset{(NM\times 1)}{\text{vec}\,(C \xi I_N)} + \underset{(NM\times 1)}{\text{vec}\,(\zeta)} \quad , \tag{6}$$

$$\text{vec}\,(\xi) = \rho_1\,(I_M \otimes B)\,\text{vec}\,(\xi) + \rho_2\,(C \otimes I_N)\,\text{vec}\,(\xi) + \text{vec}\,(\zeta) \quad . \tag{7}$$

Finalement, en changeant la notation, grâce aux équations (1) et (7), on a :

$$\underset{(R\times 1)}{Y} = \underset{(R\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(R\times 1)}{\varepsilon} \quad , \tag{8}$$

$$\varepsilon = \rho_1 W_D \varepsilon + \rho_2 W_0 \varepsilon + u \quad , \tag{9}$$

où $u \sim N(0, \sigma_u^2 I_R)$, $\varepsilon = \text{vec}(\xi)$ et $u = \text{vec}(\zeta)$ sont des vecteurs de longueurs $(R\times 1)$ et $W_D = (I_M \otimes B)$ et $W_0 = (C \otimes I_N)$ des matrices de dimension $(R\times R)$.

Le travail de spécification et d'estimation, étant relégués à des recherches subséquentes, on se limitera ici, à présenter, au prochain chapitre, la solution algorithmique d'une telle formulation.

CHAPITRE IX

ALGORITHME D'ESTIMATION D'UNE FORME AUTORÉGRESSIVE RÉSIDUELLE POUR DES DONNÉES DE FLUX

Pour des données de flux, le modèle autorégressif résiduel peut s'écrire :

(1) $Y = X\beta + \varepsilon$,

(2) $\varepsilon = \rho_1 W_D \varepsilon + \rho_2 W_O \varepsilon + u$,

$u \sim N(0, \sigma_u^2 I_R)$, avec $\varepsilon = vec(\xi)$, $u = vec(\zeta)$.

(3) $W_D = (I_M \otimes B)$,

(4) $W_O = (C \otimes I_N)$,

où $B = B(\theta)$ est la matrice de contiguïté des N zones en destination et $C = C(\theta)$ la matrice de contiguïté des M zones en origine et où l'on possède $R = M \times N$ observations. Le vecteur θ est formé de H éléments $\theta' = (\theta_1, \ldots, \theta_H)$.

Comme ce problème a la même forme qu'un SAR(2) résiduel, l'algorithme de solution est alors évident. En effet, il s'agit d'adapter les formules du chapitre IV à notre cas.

1) Fonction de vraisemblance

L'équation (2) peut s'écrire :

(5) $[I_R - \rho_1 W_D - \rho_2 W_O] \varepsilon = u$,

(6) $P(\rho_1, \rho_2, \theta) \varepsilon = u$,

où, pour simplifier la notation, on dira :

(7) $P\varepsilon = u$ et $\varepsilon = P^{-1} u$,

(8) $J(u \to \varepsilon) = |P|$,

et donc :

$$(9) \qquad V(\varepsilon) = \sigma_u^2 \, P^{-1} \, P^{-1\prime} = \sigma_u^2 \, V \quad .$$

Posons :

$$(10) \qquad \gamma' = (\rho_1, \rho_2, \theta_1, \ldots, \theta_H)$$

et $\sigma_u^2 = \sigma^2$. Alors :

$$(11) \qquad \mathcal{L}(\beta, \sigma^2, \gamma; y) = (2\pi \sigma^2)^{-R/2} \, |P| \exp\{-\frac{(y - X\beta)' \, V^{-1} \, (y - X\beta)}{2\sigma^2}\} \quad .$$

2) Dérivées premières

Comme établi à la section 2) du chapitre IV, le problème se ramène à maximiser la fonction concentrée $L^{**}(\gamma; y, \beta^*, \sigma^{*2})$, par rapport aux γ_i, c'est-à-dire :

$$(12) \qquad L^{**}(\gamma; y, \beta^*, \sigma^{*2}) = -\frac{R}{2}[\ln(2\pi) + 1] - \frac{R}{2} \ln \sigma^{*2} + \ln |P| \quad ,$$

avec

$$(13) \qquad \beta^* = (X' \, V^{-1} \, X)^{-1} \, X' \, V^{-1} \, y \quad ,$$

$$(14) \qquad \sigma^{*2} = \frac{(y - X\beta^*)' \, V^{-1} \, (y - X\beta^*)}{R} \quad ,$$

$$(15) \qquad \frac{\partial L^{**}}{\partial \gamma_i} = \mathrm{tr}\,(P^{-1} \frac{\partial P}{\partial \gamma_i}) - \frac{1}{\sigma^{*2}} (y - X\beta^*)' \, P' \, \frac{\partial P}{\partial \gamma_i} (y - X\beta^*) \quad ,$$

où

$$
(15.1)\qquad \gamma_i = \begin{array}{c|c} & \dfrac{\partial P}{\partial \gamma_i} \\ \hline \rho_1 & -(I_M \otimes B) \\ \rho_2 & -(C \otimes I_N) \\ \theta_h & -\left[\rho_1 \left(I_M \otimes \dfrac{\partial B}{\partial \theta_h}\right) + \rho_2 \left(\dfrac{\partial C}{\partial \theta_h} \otimes I_N\right)\right] \end{array} .
$$

3) Dérivées secondes

Pour les dérivées secondes, les formules sont :

$$
(16)\qquad \frac{\partial^2 L}{\partial \beta \partial \beta'} = -\frac{X' V^{-1} X}{\sigma^2} ,
$$

$$
(17)\qquad \frac{\partial^2 L}{\partial \beta \partial \sigma^2} = -\frac{X' V^{-1} \varepsilon}{\sigma^2} ,
$$

$$
(18)\qquad \frac{\partial^2 L}{\partial \beta \partial \gamma_i} = \frac{X'}{\sigma^2}\left[P' \frac{\partial P}{\partial \gamma_i} + \frac{\partial P'}{\partial \gamma_i} P\right] \varepsilon ,
$$

$$
(19)\qquad \frac{\partial^2 L}{\partial \sigma^2 \partial \sigma^2} = \frac{R}{2\sigma^4} - \frac{\varepsilon' V^{-1} \varepsilon}{\sigma^6} ,
$$

$$
(20)\qquad \frac{\partial^2 L}{\partial \sigma^2 \partial \gamma_i} = \frac{u' \dfrac{\partial P}{\partial \gamma_i} \varepsilon}{\sigma^4} ,
$$

dont on a, grâce à (15.1) :

(20.1)

γ_i =	$\partial^2 L/\partial\sigma^2\partial\gamma_i$
ρ_1	$-\dfrac{u'\ (I_M \otimes B)\ \varepsilon}{\sigma^4}$
ρ_2	$-\dfrac{u'\ (C \otimes I_N)\ \varepsilon}{\sigma^4}$
θ_h	$-\dfrac{u'\ [\rho_1\ (I_M \otimes \frac{\partial B}{\partial \theta_h}) + \rho_2\ (\frac{\partial C}{\partial \theta_h} \otimes I_N)]\ \varepsilon}{\sigma^4}$

et

$$(21)\qquad \frac{\partial^2 L}{\partial\gamma_i\partial\gamma_j} = \operatorname{tr}\left[-P^{-1}\frac{\partial P}{\partial\gamma_j}P^{-1}\frac{\partial P}{\partial\gamma_i} + P^{-1}\frac{\partial^2 P}{\partial\gamma_i\partial\gamma_j}\right] - \frac{\varepsilon'}{\sigma^2}\left[\frac{\partial P'}{\partial\gamma_j}\frac{\partial P}{\partial\gamma_i} + P'\frac{\partial^2 P}{\partial\gamma_i\partial\gamma_j}\right]\varepsilon$$

où $\dfrac{\partial P}{\partial\gamma_i}$, $\dfrac{\partial P}{\partial\gamma_j}$ sont donnés par (15.1) et où :

(21.1)

γ_j =

γ_i = $\dfrac{\partial^2 P}{\partial\gamma_i\partial\gamma_j}$	ρ_1	ρ_2	θ_g
ρ_1	0	0	$-(I_M \otimes \frac{\partial B}{\partial\theta_g})$
ρ_2	0	0	$-(\frac{\partial C}{\partial\theta_g} \otimes I_N)$
θ_h	$-(I_M \otimes \frac{\partial B}{\partial\theta_h})$	$-(\frac{\partial C}{\partial\theta_h} \otimes I_N)$	$-[\rho_1\ (I_M \otimes \frac{\partial^2 B}{\partial\theta_h\partial\theta_g}) + \rho_2(\frac{\partial^2 C}{\partial\theta_h\partial\theta_g} \otimes I_N)]$

4) Formules à programmer pour les dérivées secondes

On peut remarquer que les équations (15.1), (20.1) et (21.1) comportent des produits de kronecker. Afin de minimiser l'espace utilisé par le programme pour calculer ces équations, utilisons les identités suivantes :

i) $(I_M \otimes B)\, \varepsilon = \text{vec}\,(\xi B)$ [1] ,

ii) $(C \otimes I_N)\, \varepsilon = \text{vec}\,(C\xi)$,

où $\varepsilon = \text{vec}\,(\xi)$,

iii) $$P^{-1}\,(I_M \otimes B) = \begin{bmatrix} P^{11} & P^{12} & \cdots & P^{1M} \\ P^{21} & P^{22} & \cdots & P^{2M} \\ \vdots & \vdots & & \vdots \\ P^{M1} & P^{M2} & \cdots & P^{MM} \end{bmatrix} \begin{bmatrix} B & 0 & \cdots & 0 \\ 0 & B & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & B \end{bmatrix}$$

$$= \begin{bmatrix} P^{11}B & P^{12}B & \cdots & P^{1M}B \\ P^{21}B & P^{22}B & \cdots & P^{2M}B \\ \vdots & \vdots & & \vdots \\ P^{M1}B & P^{M2}B & \cdots & P^{MM}B \end{bmatrix}$$

où B et P^{ij} sont des matrices carrées N×N et P^{ij} est la sous-matrice ij de P^{-1}.

---

[1]. $\text{vec}\,(ABC) = (A \otimes C')\,\text{vec}\,B$.

$$\text{iv)}\quad P^{-1}\,(C \otimes I_N) = \begin{bmatrix} P^{11} & P^{12} & \cdots & P^{1M} \\ P^{21} & P^{22} & \cdots & P^{2M} \\ \vdots & \vdots & & \vdots \\ P^{M1} & P^{M2} & & P^{MM} \end{bmatrix} \begin{bmatrix} C_{11}I_N & C_{12}I_N & \cdots & C_{1M}I_N \\ C_{21}I_N & C_{22}I_N & \cdots & C_{2M}I_N \\ \vdots & \vdots & & \vdots \\ C_{M1}I_N & C_{M2}I_N & \cdots & C_{MM}I_N \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{M} C_{i1} P^{1i} & \sum_{i=1}^{M} C_{i2} P^{1i} & \cdots & \sum_{i=1}^{M} C_{iM} P^{1i} \\ \sum_{i=1}^{M} C_{i1} P^{2i} & \sum_{i=1}^{M} C_{i2} P^{2i} & \cdots & \sum_{i=1}^{M} C_{iM} P^{2i} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{M} C_{i1} P^{Mi} & \sum_{i=1}^{M} C_{i2} P^{Mi} & \cdots & \sum_{i=1}^{M} C_{iM} P^{Mi} \end{bmatrix}$$

Soit : $\underset{(N\times N)}{A} = \sum_{i=1}^{M} \underset{(N\times N)}{P^{ii}} = A'$ et $\underset{(M\times M)}{E} = E'$, la matrice dont l'élément $E_{ij} = \operatorname{tr}\left[\underset{(N\times N)}{P^{ij}}\right]$. En se servant de i) à iv), les équations (15) et (15.1) permettent d'écrire :

$$\frac{\partial L^{**}}{\partial \rho_1} = -\operatorname{tr}(A\,B) + \frac{1}{\sigma^2}\, u' \operatorname{vec}(\xi B) \quad ,$$

$$\frac{\partial L^{**}}{\partial \rho_2} = -\operatorname{tr}(CE) + \frac{1}{\sigma^2}\, u' \operatorname{vec}(C\xi) \quad ,$$

$$\frac{\partial L^{**}}{\partial \theta_h} = -\left[\rho_1 \operatorname{tr}\left(\xi \frac{\partial B}{\partial \theta_h}\right) + \rho_2 \operatorname{tr}\left(E \frac{\partial C}{\partial \theta_h}\right)\right] + \frac{\rho_1\, u'}{\sigma^2} \operatorname{vec}\left(\xi \frac{\partial B}{\partial \theta_h}\right) + \frac{\rho_2\, u'}{\sigma^2} \operatorname{vec}\left(\frac{\partial C}{\partial \theta_h} \xi\right)$$

(20.1) devient alo :

		$\frac{\partial^2 L}{\partial\sigma^2 \partial\gamma_i}$
γ_i =	ρ_1	$-\frac{u' \text{ vec } (\xi B)}{\sigma^4}$
	ρ_2	$-\frac{u' \text{ vec } (C\xi)}{\sigma^4}$
	θ_h	$-\frac{u'}{\sigma^4}\left[\rho_1 \text{ vec } (\xi \frac{\partial B}{\partial \theta_h}) + \rho_2 \text{ vec } (\frac{\partial C}{\partial \theta_h} \xi)\right]$

En se servant des identités suivantes :

v) $$\text{tr}\left[P^{-1}(I_M \otimes B)\, P^{-1}(I_M \otimes B)\right] = \text{tr}\begin{bmatrix} P^{11}B & \cdots & P^{1M}B \\ \vdots & & \vdots \\ P^{M1}B & \cdots & P^{MM}B \end{bmatrix}\begin{bmatrix} P^{11}B & \cdots & P^{1M}B \\ \vdots & & \vdots \\ P^{M1}B & \cdots & P^{MM}B \end{bmatrix}$$

$$= \text{tr}\left[\sum_{i=1}^{M}\sum_{j=1}^{M} P^{ij} B\, P^{ji} B\right] ,$$

vi) $\operatorname{tr}\,[P^{-1}\,(C \otimes I_N)\,P^{-1}\,(I_M \otimes B)]$

$$= \operatorname{tr}\begin{bmatrix} \sum_{i=1}^{M} C_{il}\,P^{li} & \cdots & \sum_{i=1}^{M} C_{iM}\,P^{li} \\ \vdots & & \vdots \\ \sum_{i=1}^{M} C_{il}\,P^{Mi} & \cdots & \sum_{i=1}^{M} C_{iM}\,P^{Mi} \end{bmatrix} \begin{bmatrix} P^{11}B & \cdots & P^{1M}B \\ \vdots & & \vdots \\ P^{M1}B & \cdots & P^{MM}B \end{bmatrix}$$

$$= \operatorname{tr}\left[\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{i=1}^{M} C_{ij}\,P^{ki}\,P^{jk}\,B\right] \quad ,$$

vii) $\operatorname{tr}\,[P^{-1}\,(C \otimes I_N)\,P^{-1}\,(C \otimes I_N)]$

$$= \operatorname{tr}\begin{bmatrix} \sum_{i=1}^{M} C_{il}\,P^{li} & \cdots & \sum_{i=1}^{M} C_{iM}\,P^{li} \\ \vdots & & \vdots \\ \sum_{i=1}^{M} C_{il}\,P^{Mi} & \cdots & \sum_{i=1}^{M} C_{iM}\,P^{Mi} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{M} C_{il}\,P^{li} & \cdots & \sum_{i=1}^{M} C_{iM}\,P^{li} \\ \vdots & & \vdots \\ \sum_{i=1}^{M} C_{il}\,P^{Mi} & \cdots & \sum_{i=1}^{M} C_{iM}\,P^{Mi} \end{bmatrix}$$

$$= \operatorname{tr}\left[\sum_{\ell}^{M}\sum_{k}^{M}\sum_{i}^{M}\sum_{j}^{M} C_{ik}C_{j\ell}\,P^{\ell i}\,P^{kj}\right] \quad .$$

Pour $\gamma_i = \rho_1, \rho_2, \theta_h$ $(h = 1,\ldots,H)$, et θ_g $(g = 1,\ldots,G)$, l'expression $\dfrac{\partial^2 L}{\partial\gamma_i\partial\gamma_j}$ devient, par v) et i) :

$$\frac{\partial^2 L}{\partial\rho_1^2} = -\operatorname{tr}\left[\sum_{i}^{M}\sum_{j}^{M} P^{ij}\,B\,P^{ji}\,B\right] - \frac{\operatorname{vec}(\xi B)'\,\operatorname{vec}(\xi B)}{\sigma^2} \quad .$$

De la même façon, par vi), i) et ii), on a :

$$\frac{\partial^2 L}{\partial \rho_1 \partial \rho_2} = - \operatorname{tr} \left[\sum_{j=1}^{M} \sum_{k=1}^{M} \sum_{i=1}^{M} C_{ij} \, P^{ki} \, P^{jk} \, B \right] - \frac{\operatorname{vec} (C\xi)' \operatorname{vec} (\xi B)}{\sigma^2}$$

et par vii) et ii),

$$\frac{\partial^2 L}{\partial \rho_2^2} = - \operatorname{tr} \left[\sum_{\ell}^{M} \sum_{k}^{M} \sum_{i}^{M} \sum_{j}^{M} C_{ik} \, C_{j\ell} \, P^{\ell i} \, P^{kj} \right] - \frac{\operatorname{vec} (C\xi)' \operatorname{vec} (C\xi)}{\sigma^2} \quad .$$

En appliquant les formules i) à vii), on a :

$$\frac{\partial^2 L}{\partial \theta_g \partial \rho_1} = - \rho_1 \operatorname{tr} \left[\sum_{i}^{M} \sum_{j}^{M} P^{ij} \, B \, P^{ji} \, \frac{\partial B}{\partial \theta_g} \right] - \rho_2 \operatorname{tr} \left[\sum_{j}^{M} \sum_{k}^{M} \sum_{i}^{M} \frac{\partial C_{ij}}{\partial \theta_g} \, P^{ki} \, P^{jk} \, B \right]$$

$$- \operatorname{tr} \left(A \, \frac{\partial B}{\partial \theta_g} \right) - \frac{\rho_1}{\sigma^2} \operatorname{vec} (\xi B)' \operatorname{vec} \left(\xi \, \frac{\partial B}{\partial \theta_g} \right) - \frac{\rho_2}{\sigma^2} \operatorname{vec} (\xi B)' \operatorname{vec} \left(\frac{\partial C}{\partial \theta_g} \, \xi \right)$$

$$+ \frac{u'}{\sigma^2} \operatorname{vec} \left(\xi \, \frac{\partial B}{\partial \theta_g} \right) \quad ,$$

$$\frac{\partial^2 L}{\partial \theta_g \partial \rho_2} = - \rho_1 \operatorname{tr} \left[\sum_{j}^{M} \sum_{k}^{M} \sum_{i}^{M} C_{ij} \, P^{ki} \, P^{jk} \, \frac{\partial B}{\partial \theta_g} \right] - \rho_2 \operatorname{tr} \left[\sum_{\ell}^{M} \sum_{k}^{M} \sum_{i}^{M} \sum_{j}^{M} C_{ik} \, \frac{\partial C_{j\ell}}{\partial \theta_g} \, P^{\ell i} \, P^{kj} \right]$$

$$- \operatorname{tr} \left[E \, \frac{\partial C}{\partial \theta_g} \right] - \frac{\rho_1}{\sigma^2} \operatorname{vec} (C\xi)' \operatorname{vec} \left(\xi \, \frac{\partial B}{\partial \theta_g} \right) - \frac{\rho_2}{\sigma^2} \operatorname{vec} (C\xi)' \operatorname{vec} \left(\frac{\partial C}{\partial \theta_g} \, \xi \right)$$

$$+ \frac{u'}{\sigma^2} \operatorname{vec} \left(\frac{\partial C}{\partial \theta_g} \, \xi \right) \quad ,$$

$$\frac{\partial^2 L}{\partial\theta_g \partial\theta_h} = -\rho_1^2 \operatorname{tr}\left[\sum_i^M \sum_j^M P^{ij} \frac{\partial B}{\partial\theta_h} P^{ji} \frac{\partial B}{\partial\theta_g}\right] - \rho_1 \rho_2 \operatorname{tr}\left[\sum_j^M \sum_k^M \sum_i^M \frac{\partial C_{ij}}{\partial\theta_g} P^{ki} P^{jk} \frac{\partial B}{\partial\theta_h}\right]$$

$$- \rho_1 \rho_2 \operatorname{tr}\left[\sum_j^M \sum_k^M \sum_i^M \frac{\partial C_{ij}}{\partial\theta_h} P^{ki} P^{jk} \frac{\partial B}{\partial\theta_g}\right] - \rho_2^2 \operatorname{tr}\left[\sum_\ell^M \sum_k^M \sum_i^M \sum_j^M \frac{\partial C_{ik}}{\partial\theta_h} \frac{\partial C_{j\ell}}{\partial\theta_g} P^{\ell i} P^{kj}\right]$$

$$- \rho_1 \operatorname{tr}\left[A \frac{\partial^2 B}{\partial\theta_g \partial\theta_h}\right] - \rho_2 \operatorname{tr}\left(E \frac{\partial^2 C}{\partial\theta_g \partial\theta_h}\right)$$

$$- \frac{\rho_1^2}{\sigma^2} \operatorname{vec}\left(\xi \frac{\partial B}{\partial\theta_h}\right)' \operatorname{vec}\left(\xi \frac{\partial B}{\partial\theta_g}\right) - \frac{\rho_1\rho_2}{\sigma^2} \operatorname{vec}\left(\xi \frac{\partial B}{\partial\theta_h}\right)' \operatorname{vec}\left(\frac{\partial C}{\partial\theta_g} \xi\right)$$

$$- \frac{\rho_1\rho_2}{\sigma^2} \operatorname{vec}\left(\frac{\partial C}{\partial\theta_h} \xi\right)' \operatorname{vec}\left(\xi \frac{\partial B}{\partial\theta_g}\right) - \frac{\rho_2^2}{\sigma^2} \operatorname{vec}\left(\frac{\partial C}{\partial\theta_h} \xi\right)' \operatorname{vec}\left(\frac{\partial C}{\partial\theta_g} \xi\right)$$

$$+ \frac{\rho_1}{\sigma^2} u' \operatorname{vec}\left(\xi \frac{\partial^2 B}{\partial\theta_g \partial\theta_h}\right) + \frac{\rho_2}{\sigma^2} u' \operatorname{vec}\left(\frac{\partial^2 C}{\partial\theta_g \partial\theta_h} \xi\right) .$$

CHAPITRE X

LES MOINDRES CARRÉS GÉNÉRALISÉS (MCG)
POUR UN PROBLÈME HÉTÉROSPATIAL

Dans les chapitres précédents, afin de tenir compte de la corrélation spatiale dans les résidus $E(\varepsilon\varepsilon') = \sigma^2 V \neq \sigma^2 I$, nous avons postulé une forme SAR(1) telle que $\varepsilon = [I - \rho W(\theta)]^{-1} u$, noté $\varepsilon = P^{-1}(\theta)\, u \sim N(0, \sigma^2 P(\theta)^{-1} P(\theta)^{-1\prime})$ et $V = P^{-1}(\theta)\, P^{-1}(\theta)'$. De cette façon, l'estimation de la matrice $V(\theta)$ s'est faite par l'entremise de l'estimation de la matrice $P(\theta)$. Dans le présent chapitre et les chapitres subséquents, nous adoptons une approche différente. Au lieu de postuler une forme fonctionnelle pour les éléments de la matrice $P(\theta)$, nous nous intéressons à estimer directement les éléments de la matrice $V(\theta)$. Pour tenir compte de cette différence, nous changeons la notation et posons $E(\varepsilon\varepsilon') = \sigma^2 \Omega(\theta)$.

1) <u>Décomposition en deux étapes</u>

Soit le modèle de régression linéaire suivant :

$$(1) \qquad \underset{(N\times 1)}{y} = \underset{(N\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(N\times 1)}{\varepsilon}$$

où on a N = nombre de régions dans l'étude

K = nombre de variables explicatives

$\underset{(N\times K)}{X}$ matrice non stochastique

$$E(\varepsilon) = \underset{(N\times 1)}{0}$$

$$E(\varepsilon\varepsilon') = \sigma^2 \Omega = E \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \cdots\cdots & \varepsilon_1\varepsilon_N \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \cdots\cdots & \varepsilon_2\varepsilon_N \\ \vdots & \vdots & & \vdots \\ \varepsilon_N\varepsilon_1 & \varepsilon_N\varepsilon_2 & \cdots\cdots & \varepsilon_N^2 \end{bmatrix}$$

Posons comme forme fonctionnelle qui modélise l'hétéroscédasticité la forme multiplicative suivante, que l'on considère pour l'instant connue :

$$(2) \qquad \sigma_1^2 = E(\varepsilon_1^2) = \sigma^2 P_1^\gamma \ldots\ldots \sigma_N^2 = E(\varepsilon_N^2) = \sigma^2 P_N^\gamma \quad .$$

Alors :

$$(3) \qquad E(\varepsilon\varepsilon') = \sigma^2 \begin{bmatrix} P_1^\gamma & \omega_{12} & \ldots\ldots & \omega_{1N} \\ \omega_{21} & P_2^\gamma & \ldots\ldots & \omega_{2N} \\ \vdots & \vdots & & \vdots \\ \omega_{N1} & \omega_{N2} & \ldots\ldots & P_N^\gamma \end{bmatrix}$$

s'écrit encore :

$$(4) \qquad E(\varepsilon\varepsilon') = \sigma^2 \begin{bmatrix} P_1^{\gamma/2} & 0 & \ldots\ldots & 0 \\ 0 & P_2^{\gamma/2} & \ldots\ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots\ldots & P_N^{\gamma/2} \end{bmatrix}$$

$$\cdot \begin{bmatrix} 1 & \frac{\omega_{12}}{P_1^{\gamma/2} P_2^{\gamma/2}} & \ldots\ldots & \frac{\omega_{1N}}{P_1^{\gamma/2} P_N^{\gamma/2}} \\ \frac{\omega_{21}}{P_2^{\gamma/2} P_1^{\gamma/2}} & 1 & \ldots\ldots & \frac{\omega_{2N}}{P_2^{\gamma/2} P_N^{\gamma/2}} \\ \vdots & \vdots & & \vdots \\ \frac{\omega_{N1}}{P_N^{\gamma/2} P_1^{\gamma/2}} & & \ldots\ldots & 1 \end{bmatrix} \cdot \begin{bmatrix} P_1^{\gamma/2} & 0 & \ldots & 0 \\ 0 & P_2^{\gamma/2} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & P_N^{\gamma/2} \end{bmatrix}$$

On retrouve au centre de l'expression à droite de l'égalité, la matrice des corrélations résiduelles, que l'on nomme R. L'équation (4) devient alors :

$$(5) \qquad E(\varepsilon\varepsilon') = \sigma^2 \Omega = \sigma^2 \cdot D \cdot R \cdot D \quad ,$$

où $\underset{(N\times N)}{D} = \text{diag}\ (P_1^{\gamma/2}\ P_2^{\gamma/2}\ \ldots\ P_N^{\gamma/2})$. On a donc :

$$(6) \quad (\sigma^2\Omega)^{-1} = \sigma^{-2}\Omega^{-1} = \sigma^{-2}\ \text{diag}(P_1^{-\gamma/2}\ P_2^{-\gamma/2}\ \ldots P_N^{-\gamma/2})\ R^{-1}\ \text{diag}(P_1^{-\gamma/2}\ P_2^{-\gamma/2}\ \ldots P_N^{-\gamma/2})$$

Donc, pour le modèle linéaire (1), le BLUE (meilleur estimateur linéaire non biaisé) de β s'écrit :

$$(7) \qquad \hat{\beta} = (X'\ \Omega^{-1}\ X)^{-1}\ X'\ \Omega^{-1}\ y$$

$$= (X'\ D^{-1}\ R^{-1}\ D^{-1}\ X)^{-1}\ X'\ D^{-1}\ R^{-1}\ D^{-1}\ y \quad ,$$

$$(8) \qquad \hat{\beta} = (X^{*\prime}\ R^{-1}\ X^{*})^{-1}\ X^{*\prime}\ R^{-1}\ y^{*}$$

et

$$(9) \qquad V(\hat{\beta}) = \sigma^2\ (X^{*\prime}\ R^{-1}\ X^{*})^{-1} \quad ,$$

où $X^{*} = \text{diag}\ (P_1^{-\gamma/2}\ P_2^{-\gamma/2}\ \ldots\ P_N^{-\gamma/2})\ X$ et $y^{*} = \text{diag}\ (P_1^{-\gamma/2}\ P_2^{-\gamma/2}\ \ldots\ P_N^{-\gamma/2})\ y$.

Pour des matrices D et R connues, la décomposition utilisée pour former les équations (8) et (9) montre que l'on peut obtenir la solution BLUE de β:

$$\hat{\beta} = (X'\ \Omega^{-1}\ X)^{-1}\ X'\ \Omega^{-1}\ y \quad ,$$

par une *procédure composée de 2 étapes*. Dans une première, on transforme le modèle original (1) en divisant chaque observation i par la fonction $P_i^{\gamma/2}$. C'est la procédure habituellement employée pour corriger l'hétéroscédasticité. On peut alors former le modèle homoscédastique suivant :

$$(10) \qquad y^* = X^* \beta + \varepsilon^* \quad ,$$

où $E(\varepsilon^*) = 0$,

$E(\varepsilon^* \varepsilon^{*\prime}) = \sigma^2 R$.

A la deuxième étape, il suffit d'appliquer l'estimateur MCG (8) au modèle transformé (10) :

$$\hat{\beta} = (X^{*\prime} R^{-1} X^*)^{-1} X^{*\prime} R^{-1} y^* \quad .$$

Comme on traite des données spatiales (de coupe transversale), on nomme la matrice R la matrice des corrélations spatiales.

Comme c'est habituellement le cas, les éléments qui forment les matrices D et R sont inconnus. Lorsqu'on estime d'une façon quelconque les paramètres qui forment ces matrices, on parle des MCGE (moindres carrés généralisées estimés) que l'on note :

$$(11) \qquad \hat{\hat{\beta}} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y \quad .$$

2) Quelques propriétés des estimateurs moindres carrés généralisés

2.1) Ω connu

Pour Ω connu, l'estimateur BLUE de β est $\hat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y$. En effet, selon nos hypothèses, comme X et Ω sont des matrices fixes, donc non stochastiques, $\hat{\beta}$ est une combinaison linéaire du vecteur aléatoire y. Dans les cours de base d'économétrie, on vérifie que $\hat{\beta}$ est le meilleur estimateur linéaire non biaisé (BLUE). C'est donc dire qu'aucun autre estimateur linéaire non biaisé possède une matrice variance-covariance "plus petite" que celle de $\hat{\beta} = \sigma^2 (X' \Omega^{-1} X)^{-1} = \sum_{\hat{\beta}}$. Soit $\bar{\beta}$ tout autre estimateur non biaisé, la soustraction des matrices de variance et covariance suivantes $\sum_{\bar{\beta}} - \sum_{\hat{\beta}}$ donne une matrice semi-définie positive[1]. C'est ce qu'implique le théorème Gauss-Markov. Pour obtenir ce résultat, il suffit d'utiliser l'espérance mathématique, concept employé lors de l'étude des propriétés d'estimateurs en échantillon fini.

Lorsqu'on remplace Ω par une matrice estimée $\hat{\Omega}$, $\hat{\hat{\beta}} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y$ possède des propriétés d'échantillon fini qui s'avèrent plus difficiles à évaluer. Ceci s'explique par le fait que $\hat{\hat{\beta}}$ devient alors une fonction complexe de variables aléatoires présentes dans y et $\hat{\Omega}$. Le problème devient beaucoup moins ardu lorsqu'on étudie les propriétés asymptotiques. On emploie alors les concepts de convergence, de limite en probabilité et de loi de distribution asymptotique. Pour un rappel, consulter l'appendice III.

---

1. Voir Judge, Hill, Griffiths, Lütkepohl et Lee, 1982, Introduction to the Theory and Practice of Econometrics, John Wiley & Sons, Chapter 10.

Pourvu qu'on ajoute à notre modèle les hypothèses suivantes :

$$(12) \qquad \lim_{N\to\infty} \frac{X' \Omega^{-1} X}{N} = Q$$

où Q est définie positive, le vecteur d'erreur ε est de moyenne zéro et $\forall N$ possède une matrice variance-covariance $\sigma^2 \underset{(N\times N)}{\Omega}$ non singulière, pour

$$(13) \qquad \hat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \quad ,$$

qui s'écrit :

$$(14) \qquad \hat{\beta} = \beta + (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \varepsilon \quad ,$$

les conditions $E(\hat{\beta}) = \beta$ et grâce à (12),

$$(15) \qquad \lim_{N\to\infty} \frac{1}{N} \left(\frac{X' \Omega^{-1} X}{N}\right)^{-1} = \lim_{N\to\infty} \frac{1}{N} \cdot Q^{-1} = 0 \quad ,$$

sont suffisantes pour vérifier la convergence de $\hat{\beta}$, c'est-à-dire plim $\hat{\beta} = \beta$ [1]. Une application du théorème central limite permet d'établir sous des conditions supplémentaires que :

$$(16) \qquad \sqrt{N}\,(\hat{\beta} - \beta) = \left(\frac{X' \Omega^{-1} X}{N}\right)^{-1} \frac{X' \Omega^{-1} \varepsilon}{\sqrt{N}}$$

possède une loi de distribution asymptotique normale, c'est-à-dire[2] :

$$(17) \qquad \sqrt{N}\,(\hat{\beta} - \beta) \sim N\left(0,\ \sigma^2 \lim_{N\to\infty} \left(\frac{X' \Omega^{-1} X}{N}\right)^{-1}\right) \quad .$$

---

1. Dhrymes (1978), p. 102.

2. Theil (1971), section 8.6.

En effet, si toutes les conditions sont respectées :

$$(18) \qquad \frac{X' \Omega^{-1} \varepsilon}{\sqrt{N}} \xrightarrow{D} z \sim N\left(0, \lim_{N\to\infty} \sigma^2 \cdot \frac{X' \Omega^{-1} X}{N}\right)$$

$$z \sim N (0, \sigma^2 Q) \quad .$$

Comme par (12) le premier terme de droite de l'équation (16) a une limite Q^{-1}, par l'application de la règle 8 de l'appendice III, $\sqrt{N} (\hat{\beta} - \beta) \sim N (0, \sigma^2 Q^{-1})$.

2.2) Ω inconnu

Sous des conditions additionnelles, s'il est possible de trouver une estimation convergente $\hat{\Omega}$ de Ω, alors dans ce cas, $\sqrt{N} (\hat{\beta} - \beta)$ et $\sqrt{N} (\hat{\hat{\beta}} - \beta)$ auront *la même loi de distribution asymptotique*. Ces conditions sont :

$$(19) \qquad \operatorname*{plim}_{N\to\infty} \frac{1}{N} X' (\hat{\Omega}^{-1} - \Omega^{-1}) X = 0 \quad ,$$

$$(20) \qquad \operatorname*{plim}_{N\to\infty} \frac{1}{\sqrt{N}} X' (\hat{\Omega}^{-1} - \Omega^{-1}) \varepsilon = 0 \quad .$$

Si de plus :

$$(21) \qquad \operatorname{plim} \frac{1}{N} \varepsilon' (\hat{\Omega}^{-1} - \Omega^{-1}) \varepsilon = 0 \quad ,$$

alors :

$$\hat{\hat{\sigma}}^2 = \frac{(y - X\hat{\hat{\beta}})' \hat{\Omega}^{-1} (y - X\hat{\hat{\beta}})}{N-K} \quad ,$$

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})' \Omega^{-1} (y - X\hat{\beta})}{N-K} \quad ,$$

sont des estimateurs convergents de σ^2 [1]. En effet, par (18), (20) et la règle 7 de l'appendice III :

$$(22) \qquad \frac{X' \hat{\Omega}^{-1} \varepsilon}{\sqrt{N}} \xrightarrow{D} z \sim N(0, \sigma^2 Q) \quad ,$$

alors, par (19) $\underset{N\to\infty}{\text{plim}} \frac{X' \hat{\Omega}^{-1} X}{N} = Q$ et par la règle 8 de l'appendice III, on obtient le résultat désiré, soit : $\sqrt{N}(\hat{\beta} - \beta)$ et $\sqrt{N}(\hat{\hat{\beta}} - \beta)$ ont la même loi de distribution limite.

Alors, si ces conditions sont respectées et si on possède une matrice $\hat{P}$ telle que $\hat{P}' \hat{P} = \hat{\Omega}^{-1}$, la procédure usuelle appliquée au modèle transformé $\hat{P} y = \hat{P} X\beta + \hat{P} \varepsilon$ est asymptotiquement justifiée.

2.3) Propriétés de l'estimateur des moindres carrés ordinaires

En ce qui concerne l'estimateur b du moindre carré ordinaire (MCO) :

$$(23) \qquad b = \beta + (X'X)^{-1} X' \varepsilon \quad ,$$

b est convergent et suite à l'application du théorème Gauss-Markov, asymptotiquement moins efficace que le BLUE :
$\hat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y$. On a $E(b) = \beta$ et de plus :

$$\underset{N\to\infty}{\lim} \text{var}(b) = \frac{\sigma^2}{N} \left(\frac{X'X}{N}\right)^{-1} \frac{X' \Omega X}{N} \left(\frac{X'X}{N}\right)^{-1} = \underset{(K\times K)}{0} \quad ,$$

car sous les conditions supplémentaires :

---

[1]. On met les équations (19) à (21) plutôt que $\underset{N\to\infty}{\text{plim}} \hat{\Omega}^{-1} - \Omega^{-1} = 0$, car dans cette dernière expression, la taille de la matrice $\hat{\Omega}$ augmente avec celle de l'échantillon, ce qui n'est pas permis par l'opérateur $\underset{N\to\infty}{\text{plim}}$.

$$\lim_{N\to\infty} \left(\frac{X'X}{N}\right) = \Phi \qquad \text{matrice finie et définie positive ,}$$

(24)

$$\lim_{N\to\infty} \frac{X'\ \Omega\ X}{N} = \Gamma \qquad \text{matrice finie et définie positive ,}$$

$$\lim_{N\to\infty} \text{var } (b) = \lim \frac{\sigma^2}{N} \cdot \Phi^{-1}\ \Gamma\ \Phi^{-1} = 0.$$

Avec les hypothèses (24) et en ajoutant la normalité à l'erreur ε, on vérifie que la distribution asymptotique de b est non singulière.

$$(25) \qquad \sqrt{N}\ (b - \beta) \sim N\ (0,\ \sigma^2\ \Phi^{-1}\ \Gamma\ \Phi^{-1}) \ .$$

3) L'estimateur $\tilde{\beta}$ du maximum de vraisemblance

Lorsqu'on ajoute la normalité aux erreurs ε, c'est-à-dire si $\varepsilon \sim N\ (0,\ \sigma^2\ \Omega)$ avec Ω inconnu, on obtient habituellement un estimateur $\tilde{\beta}$ de la même forme que les MCGE[1], en effet, on a :

$$(26) \qquad \tilde{\beta} = (X'\ \tilde{\Omega}^{-1}\ X)^{-1}\ X'\ \tilde{\Omega}^{-1}\ y \ .$$

Dans l'expression (26), $\tilde{\Omega}$ n'est pas nécessairement égal à $\hat{\Omega}$ de (11). Toutefois, $\hat{\beta}$, $\tilde{\beta}$ et $\hat{\tilde{\beta}}$ auront asymptotiquement les mêmes propriétés. Comme on le sait, sous des conditions générales, $\tilde{\beta}$ sera habituellement convergent, asymptotiquement normal et asymptotiquement efficace. Basé sur ce résultat, $\hat{\beta}$, $\tilde{\beta}$ et $\hat{\tilde{\beta}}$ atteindront tous asymptotiquement la borne Cramer-Rao asymptotique.

---

[1]. Dans la littérature économétrique, il existe une exception à cette règle. En effet, pour un modèle hétéroscédastique où $E(\varepsilon_t^2) = \sigma^2\ E(y_t^2)$ où $y_t \sim N$, l'estimateur $\tilde{\beta}$ de β n'est pas de type MCGE (Judge (1980), p. 140).

Remarque

En statistique, les preuves de convergence, de normalité et d'efficacité asymptotique des estimateurs du maximum de vraisemblance sont établies à partir de la loi de densité jointe d'un échantillon aléatoire $\mathcal{L} = \prod^{N} f(\cdot)$, caractérisé par l'indépendance entre les observations. Pour Ω inconnu avec $\Omega = \Omega(\theta)$, où $\underset{(M\times 1)}{\theta}$ est inconnu, lorsque $\Omega^{-1}(\theta)$ peut s'écrire comme $P(\theta)'\, P(\theta)$, ces dernières propriétés optimales peuvent être étendues au vecteur $\tilde{\zeta} = (\tilde{\beta}', \tilde{\theta}', \tilde{\sigma}^2)$. Pour se convaincre, il suffit de considérer le fait que les résidus $v = P\,\varepsilon$ du modèle transformé $Py = P\,X\beta + v$ sont indépendants.

Dans un problème d'autocorrélation spatiale ou temporelle ou bien encore un problème d'hétéroscédasticité, on estime habituellement la matrice $\Omega(\theta)$ par l'intermédiaire d'une modélisation des éléments de la matrice $P(\theta)$. Dans le cas présent, contrairement à cette approche, les formes fonctionnelles que l'on postule concernent les éléments ω_{ij} de $\Omega(\theta)$. Afin de tenir compte de cette situation, Magnus (1978) a établi, pour $\varepsilon \sim N(0, \sigma^2\,\Omega)$, les conditions qu'il faut respecter pour que le vecteur $\tilde{\zeta} = (\tilde{\beta}', \tilde{\theta}', \tilde{\sigma}^2)$ soit convergent, asymptotiquement normalement distribué et asymptotiquement efficace. D'après ses résultats, lorsque certaines conditions sont respectées, $\sqrt{N}\,(\tilde{\zeta} - \zeta) \sim N(0, \lim_{N\to\infty} [\frac{\psi}{N}]^{-1})$, où ψ représente la matrice d'information. A la section finale de ce chapitre, nous tenterons de vérifier si ce résultat s'applique pour le type de modèle que l'on veut considérer.

4) Procédure d'estimation en deux étapes

Dans cette section, on va développer un algorithme d'estimation qui tiendra compte à la fois de la corrélation spatiale et de l'hétéroscédasticité. Notre but sera donc d'estimer conjointement le vecteur de paramètres β du modèle (1) et les paramètres de la matrice variance-covariance $\sigma^2 \Omega$ des résidus. Dans un tel cas, les propriétés de ces estimateurs, dans le contexte d'échantillon fini, deviennent extrêmement difficiles à étudier. C'est pourquoi on donnera à notre algorithme une justification exclusivement asymptotique.

Comme on va le voir, on exploitera la décomposition en deux étapes établie aux équations (1) à (9), c'est-à-dire, dans un premier temps, on estimera de façon convergente les paramètres de la forme qui modélise l'hétéroscédasticité. Par la suite, on partira de cette solution convergente pour s'intéresser à une estimation du maximum de vraisemblance du modèle homoscédastique résultant lequel est sujet à la corrélation spatiale résiduelle.

4.1) Modélisation convergente de l'hétéroscédasticité

Il existe dans la littérature plusieurs types de forme fonctionnelle qui permettent de modéliser l'hétéroscédasticité. Pour une synthèse des principales formes, consulter Judge (1980), chapitre 4. Les économètres utilisent généralement la forme multiplicative et c'est d'ailleurs celle que l'on va retenir ici.

4.1.1) *La forme hétéroscédastique multiplicative*

Selon notre notation à l'équation (2), on a :
$E(\varepsilon_1^2) = \sigma^2 P_1^\gamma$, $E(\varepsilon_2^2) = \sigma^2 P_2^\gamma$ .... $E(\varepsilon_N^2) = \sigma^2 P_N^\gamma$. Posons $z_{t2} = \ln P_t$, $\alpha_1 = \ln \sigma^2$ et $\alpha_2 = \gamma$, cette opération peut s'écrire :

$$E(\varepsilon_t^2) = e^{\alpha_1 + \alpha_2 z_{t2}} \quad . \tag{27}$$

Soit z_t un vecteur S×1 de variables exogènes à l'observation t et α un vecteur S×1 de paramètres inconnus, la forme généralisée de (27) s'écrit :

$$\sigma_t^2 = E(\varepsilon_t^2) = e^{z_t'\alpha} \quad , \tag{28}$$

où $z_{t1} = 1 \qquad t = 1,\ldots,N$.

La forme (28) est appelée multiplicative car elle s'exprime comme $e^{\alpha_1} \cdot e^{\alpha_2 z_{t2}} \cdot \ldots \cdot e^{\alpha_s z_{ts}}$. Vu que $E(\varepsilon_t^2) = \sigma^2 P_t^\gamma$ est un cas particulier de (28), les propriétés des estimateurs en (28) valent aussi pour cette dernière formulation. Afin d'estimer α, la méthode usuelle suggère d'exécuter la transformation suivante :

$$\varepsilon_t^2 = e^{z_t'\alpha} + [\varepsilon_t^2 - E(\varepsilon_t^2)] \quad , \tag{29}$$

$$\varepsilon_t^2 = e^{z_t'\alpha} + [\varepsilon_t^2 - \sigma_t^2] \quad .$$

Le logarithme naturel s'écrit $\ln \varepsilon_t^2 = z_t'\alpha + \ln (\varepsilon_t^2/\sigma_t^2)$, et on a donc, finalement, le modèle linéaire suivant :

$$\ln \varepsilon_t^2 = z_t'\alpha + v_t^* \quad . \tag{30}$$

Si on fait l'hypothèse de normalité pour ε_t, alors $\varepsilon_t^2/\sigma_t^2 \sim \chi^2(1)$. Ainsi v_t^* est distribué comme le logarithme naturel d'une variable aléatoire khi carrée avec 1 degré de liberté. En faisant l'hypothèse que $E(\varepsilon_t \varepsilon_r) = 0$ $\forall t \neq r$, Harvey (1976) obtient :

$$\text{a)} \quad E(v^*) = -1{,}2704\ i \quad ,$$
$$\text{b)} \quad E(v^* v^{*\prime}) = 4{,}9348\ I_N \quad , \tag{31}$$

où $i' = (1 \ \ 1 \ \ldots \ 1)_{(1\times N)}$ et où le modèle (30) s'écrit de façon matricielle :

$$q^* = Z\alpha + v^* \quad , \tag{32}$$

avec $q^* = (\ln \varepsilon_1^2 \ \ldots \ \ln \varepsilon_N^2)'$. Comme pour notre cas, $E(\varepsilon_t \varepsilon_r) \neq 0$ est plus raisonnable, on dira :

$$\text{a)} \quad E(v^*) = -1{,}2704\ i \quad ,$$
$$\text{b)} \quad \text{var}\,(v^*) = \underset{(N\times N)}{Q} \quad , \tag{33}$$

avec Q définie positive, c'est-à-dire qu'on maintient l'hypothèse de moyenne constante $E(v_t^*) = -1{,}2704$ $\forall t$ tout en ajoutant la possibilité de variances non nulles.

Comme l'équation (30) s'écrit :

$$(34) \qquad \ln \varepsilon_t^2 = \alpha_1 + z_{t2}\alpha_2 + \ldots + z_{ts}\alpha_s + v_t^* \quad , \qquad t = 1,\ldots,N,$$

avec $E(v_t^*) = -1{,}2704$. Définissons une nouvelle erreur $w_t^* = v_t^* + 1{,}2704$ tel que $E(w_t^*) = 0 \ \forall t$. Alors le modèle (34) devient :

$$(35) \qquad \ln \varepsilon_t^2 = (\alpha_1 - 1{,}2704) + z_{t2}\alpha_2 + \ldots + z_{ts}\alpha_s + w_t^* \quad ,$$

ce que l'on note :

$$(36) \qquad q_t^* = \alpha_0 + z_{t2}\alpha_2 + \ldots + z_{ts}\alpha_s + w_t^* \quad , \qquad t = 1,\ldots,N,$$

et de façon matricielle :

$$(37) \qquad q^* = Z \begin{bmatrix} \alpha_0 \\ \alpha_2 \\ \vdots \\ \alpha_s \end{bmatrix} + w^* \quad ,$$

avec $\alpha_0 = \alpha_1 - 1{,}2704$,

$$(38) \qquad E(w^*) = \underset{(N\times 1)}{0} \quad ,$$

et grâce à (33) b) :

$$(39) \qquad E(w^* w^{*\prime}) = \underset{(N\times N)}{Q} \quad .$$

En ajoutant une hypothèse sur le comportement limite de Z :

(40) $\lim_{N\to\infty} \frac{Z'Z}{N} = \Sigma_{ZZ}$, matrice finie et définie positive,

et

(41) $\lim_{N\to\infty} \frac{Z' Q Z}{N} = \tilde{\Phi}$, matrice finie et définie positive .

Comme les équations (32) et (37) permettent d'écrire :

(42) $$q^* = Z \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_s \end{bmatrix} + v^* = Z \begin{bmatrix} \alpha_0 \\ \alpha_2 \\ \vdots \\ \alpha_s \end{bmatrix} + w^* \quad ,$$

(43) $$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_s \end{bmatrix} = (Z'Z)^{-1} Z' q^* = \begin{bmatrix} \alpha_0 \\ \alpha_2 \\ \vdots \\ \alpha_s \end{bmatrix} + (Z'Z)^{-1} Z' w^* \quad .$$

Les équations (38) et (39) alliées aux résultats de la section 2.3) permettent d'affirmer que le vecteur $[\hat{\alpha}_1 \ \hat{\alpha}_2 \ \ldots \ \hat{\alpha}_s]'$ est un estimateur convergent de $[\alpha_0 \ \alpha_2 \ \ldots \ \alpha_s]'$. Nous venons donc d'établir que l'estimation MCO du modèle $q^* = Z\alpha + v^*$ produit des estimateurs $\hat{\alpha}_2,\ldots,\hat{\alpha}_s$ convergents de $\alpha_2,\ldots,\alpha_s$. La non convergence de $\hat{\alpha}_1$ pour $\alpha_1 = \ln \sigma^2$ n'est pas tragique. En effet, notre but est d'obtenir une estimation $\hat{\sigma}_t^2$ de σ_t^2 de l'équation (28). On a maintenant :

(44) $$\hat{\sigma}_t^2 = e^{z_t'\hat{\alpha}} = e^{\hat{\alpha}_1} \cdot e^{z_{t2}\hat{\alpha}_2} \cdot \ldots e^{z_{ts}\hat{\alpha}_s} \quad , \qquad t = 1,\ldots,N \ .$$

Comme $e^{\hat{\alpha}_1} = \hat{\sigma}^2$ est un facteur de proportionnalité qui disparaît lors d'une estimation MCG [voir les équations (2) à (9)] les seules estimations pertinentes sont celles de $\alpha_2,\ldots,\alpha_s$. Pour le cas particulier de notre modèle aux équations (1) et (2), $\hat{\gamma} = \hat{\alpha}_2$ est un estimateur convergent de $\gamma = \alpha_2$.

4.1.2) *Estimation de la forme hétéroscédastique multiplicative*

Afin d'estimer la forme fonctionnelle de l'hétéroscédasticité, on a proposé, à la section précédente, d'estimer par les moindres carrés ordinaires l'équation suivante :

$$\ln \varepsilon_t^2 = z_t'\alpha + v_t^* \quad , \qquad t = 1,\ldots,N.$$

Comme ε_t de l'équation (1) est inobservable, on propose de remplacer le vecteur $\underset{(N\times 1)}{\varepsilon} = y - X\beta$ par le vecteur $\underset{(N\times 1)}{\tilde{e}} = y - Xb$, où $b = (X'X)^{-1} X'y$. En effet, tel que montré à la note [1] suivante, $\tilde{e}_t \xrightarrow{D} \varepsilon_t$, $t = 1,\ldots,N$.

---

[1]. Soit $\tilde{e} = Y - Xb = My = M\varepsilon$, où $M = I - X(X'X)^{-1} X'$ et X fixe.

$\tilde{e} - \varepsilon = - X(X'X)^{-1} X'\varepsilon \rightarrow E(\tilde{e} - \varepsilon) = 0$

$$V(\tilde{e} - \varepsilon) = \sigma^2 X(X'X)^{-1} X' \Omega X(X'X)^{-1} X' \quad .$$

Grâce à l'équation (24), on a, pour chaque erreur $\tilde{e}_t$ et ε_t,

$V(\tilde{e}_t - \varepsilon_t) = \frac{\sigma^2}{N} [X_{t1},\ldots,X_{tK}] \Phi^{-1} \Gamma \Phi^{-1} \begin{bmatrix} X_{t1} \\ X_{tK} \end{bmatrix}$, ainsi

$\lim_{N\to\infty} V(\tilde{e}_t - \varepsilon_t) = 0 \quad \forall t$, donc $\text{plim}\ (\tilde{e}_t - \varepsilon_t) = 0 \quad \forall t$. Cette condition nous assure donc par la règle 7 de l'appendice III que $\tilde{e}_t$ converge en distribution vers ε_t.

On a donc le modèle suivant : $\ln \tilde{e}_t^2 = z_t'\alpha + v_t$, où $v_t = \ln (\tilde{e}_t^2/\sigma_t^2)$ ou matriciellement, avec $q_t = \ln \tilde{e}_t^2$, $t = 1,\ldots,N$:

(45) $\quad q = Z\alpha + v \quad .$

Comme les propriétés de petits échantillons du vecteur v sont alors compliquées, il reste à nous reposer sur les propriétés asymptotiques de v. Comme $\tilde{e}_t \xrightarrow{D} \varepsilon_t \quad \forall t$, alors par la règle 9 de l'appendice III, $\ln (\tilde{e}_t^2/\sigma_t^2)$ converge en distribution vers $\ln (\varepsilon_t^2/\sigma_t^2) = v_t^*$. La démarche entreprise à la section 4.1.1) nous assure alors de la convergence des estimateurs issus de cette procédure.

4.2) Modélisation de la corrélation spatiale

A la seconde étape, on se propose d'estimer le modèle homoscédastique (10) lequel est sujet à la corrélation spatiale. D'après l'équation (9), ce modèle est construit à partir de la solution convergente de l'étape précédente, ce qui est justifiable car asymptotiquement pour l'estimation de β, il est équivalent de parler de $\hat{\gamma}$ ou γ. Alors si on répète l'équation (10) en omettant les étoiles pour simplifier la notation, on a [1] :

(46) $$\underset{(N\times 1)}{y} = \underset{(N\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(N\times 1)}{\varepsilon} \quad ,$$

où $E(\varepsilon) = 0$ et $E(\varepsilon\varepsilon') = \sigma^2 R$, où R est la matrice des corrélations spatiales des résidus qui est fonction d'un certain nombre de variables pertinentes et d'un vecteur θ de L paramètres inconnus. A la section 4.2.1), on présentera

[1]. En l'absence d'hétéroscédasticité, le modèle (10) constitue alors notre modèle initial.

une procédure MCGE dont les propriétés asymptotiques restent à étudier. A la section 4.2.3), on développera une procédure d'estimation du maximum de vraisemblance pour l'ensemble des paramètres $\underset{(K\times1)}{\beta}$, $\underset{(L\times1)}{\theta}$ et σ^2.

Remarque

L'idée de modéliser directement les corrélations spatiales résiduelles n'est pas nouvelle. En effet, W.D. Fisher (1971) a proposé une estimation MCG du modèle (10) où les éléments r_{ij} de R sont remplacés par des mesures d'accessibilité. Selon lui, il est raisonnable de croire que la corrélation entre les erreurs i et j du modèle est proportionnelle à l'accessibilité entre la paire de zones concernées. Comme mesure d'accessibilité, il utilise la réciproque du temps de trajet minimum[1] entre ces deux zones. Ainsi sa matrice de corrélation est une matrice d'accessibilité R où on a $r_{ii} = 1\ \forall i$ et r_{ij} = 1/temps de trajet entre i et j.

Selon ses termes, c'est le caractère décevant des résultats d'estimation MCG obtenus qui l'ont poussé à abandonner son approche. Depuis cet article, on a eu connaissance d'aucune autre tentative d'estimation du même genre.

4.2.1) *Une procédure MCGE*

Afin d'obtenir un estimateur de type moindres carrés généralisés estimés (MCGE), il faut tout d'abord procéder à une estimation $\hat{r}_{ij}$ des éléments r_{ij} qui forment la matrice de corrélation spatiale R, pour, par la suite, produire l'estimateur :

---

[1]. Quel que soit le mode de transport.

$$(47) \qquad \hat{\beta} = (X' \hat{R}^{-1} X)^{-1} X' \hat{R}^{-1} y \quad .$$

Comme on a :

$$(48) \qquad r_{ij} = E\left(\frac{\varepsilon_i \varepsilon_j}{\sigma^2}\right) \quad ,$$

en s'inspirant de la procédure usuelle employée dans les travaux d'hétéroscédasticité, on peut former la régression suivante qui, à cause de la symétrie de R se limite aux termes i > j

$$(49) \qquad \frac{\varepsilon_i \varepsilon_j}{\sigma^2} = E\left(\frac{\varepsilon_i \varepsilon_j}{\sigma^2}\right) + \left[\frac{\varepsilon_i \varepsilon_j}{\sigma^2} - E\left(\frac{\varepsilon_i \varepsilon_j}{\sigma^2}\right)\right] \quad ,$$

ce qui s'écrit encore :

$$(50) \qquad \frac{\varepsilon_i \varepsilon_j}{\sigma^2} = r_{ij} + \left[\frac{\varepsilon_i \varepsilon_j}{\sigma^2} - r_{ij}\right] , \qquad \text{pour } i > j \ .$$

Pour r_{ij} fonction quelconque de variables $Z^1_{ij},\ldots,Z^H_{ij}$ et de paramètres $\theta_1,\ldots,\theta_L$, on peut écrire, en posant $u_{ij} = \left[\frac{\varepsilon_i \varepsilon_j}{\sigma^2} - r_{ij}\right]$:

$$(51) \qquad \frac{\varepsilon_i \varepsilon_j}{\sigma^2} = f\left(Z^1_{ij},\ldots,Z^H_{ij}; \theta_1,\ldots,\theta_L\right) + u_{ij} \quad , \qquad \text{pour } i > j \ .$$

Cette formulation (51) possède pour l'instant de nombreuses lacunes. L'étude de la loi de distribution de u_{ij} de moyenne nulle exige un peu de travail et pour des raisons que l'on verra, nous n'avons pas jugé, ici, pertinent de le faire. Deuxièmement, comme le vecteur d'erreur ε est inobservable, on peut le remplacer par son substitut naturel, le vecteur d'erreur résiduelle

$\tilde{e} = y - Xb$, car comme on l'a déjà vu, $\tilde{e}_t \overset{D}{\to} \varepsilon_t \ \forall t$. De cette façon les propriétés de cette méthodologie deviennent attirantes principalement dans le contexte d'échantillon infini[1]. Finalement, la matrice $\hat{R}$, formée par les éléments estimés $\hat{r}_{ij}$ issus de la régression (50) doit être, comme on le sait, définie positive. Pour une forme non contrainte telle qu'en (51), la probabilité que $\hat{R}$ soit définie positive est assez faible.

A cause de ces raisons, comme on n'est pas en position d'évaluer toutes les propriétés de l'estimation $\hat{r}_{ij}$ de r_{ij}, on proposera à la section 4.2.3) une méthode qui généralement produit des estimateurs convergents. On parle ici, d'une procédure du maximum de vraisemblance.

Mais avant, étudions plus spécifiquement la fonction $r_{ij} = f(Z^1_{ij}, \ldots, Z^H_{ij}, \theta_1, \ldots, \theta_L)$. A la prochaine section, on va déterminer une forme fonctionnelle précise qui permettra de modéliser de façon raisonnable la corrélation spatiale.

4.2.2) *Une forme fonctionnelle raisonnable pour la modélisation de la corrélation spatiale*

Par définition, une corrélation se situe dans l'intervalle $-1 \leq r_{ij} \leq 1$. Il serait alors intéressant de choisir une forme fonctionnelle qui respecte cet intervalle. Comme forme possible, on propose :

---

1. Comme $E\left(\frac{\tilde{e}\tilde{e}'}{\sigma^2}\right) = MRM$, où $M = I - X(X'X)^{-1}X'$, $\hat{u}_{ij} = \left[\frac{\tilde{\varepsilon}_i \tilde{\varepsilon}_j}{\sigma^2} - r_{ij}\right]$ sera généralement de moyenne non nulle et non constante.

$$
(52)\qquad
\begin{aligned}
r_{ij} &= 1 \qquad , \qquad i = j \\
r_{ij} &= \frac{1 - e^{-Z'_{ij}\alpha}}{1 + e^{-Z'_{ij}\alpha}} \, , \qquad i > j
\end{aligned}
$$

où $\underset{(H\times 1)}{Z_{ij}} = [Z^1_{ij},\ldots,Z^H_{ij}]'$ et où $\underset{(H\times 1)}{\alpha} = [\alpha_1,\ldots,\alpha_H]'$.

A la figure 1, on rappelle le comportement de la fonction e^{-x}. C'est à la figure 2 que l'on étudie celui de $\frac{1-e^{-x}}{1+e^{-x}}$.

FIGURE 1

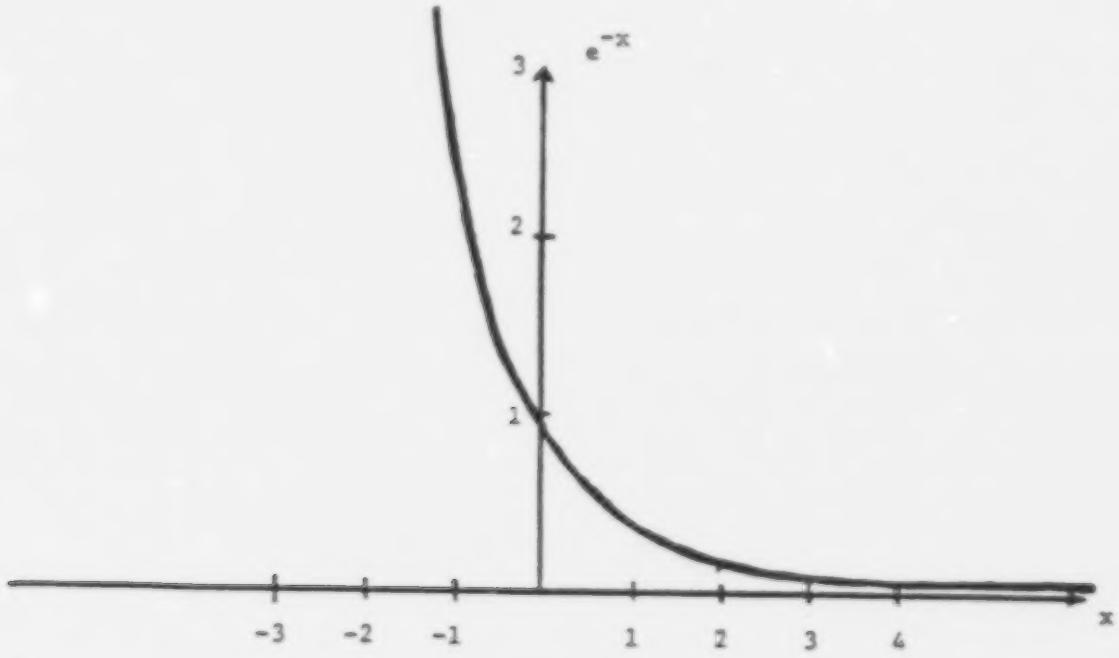
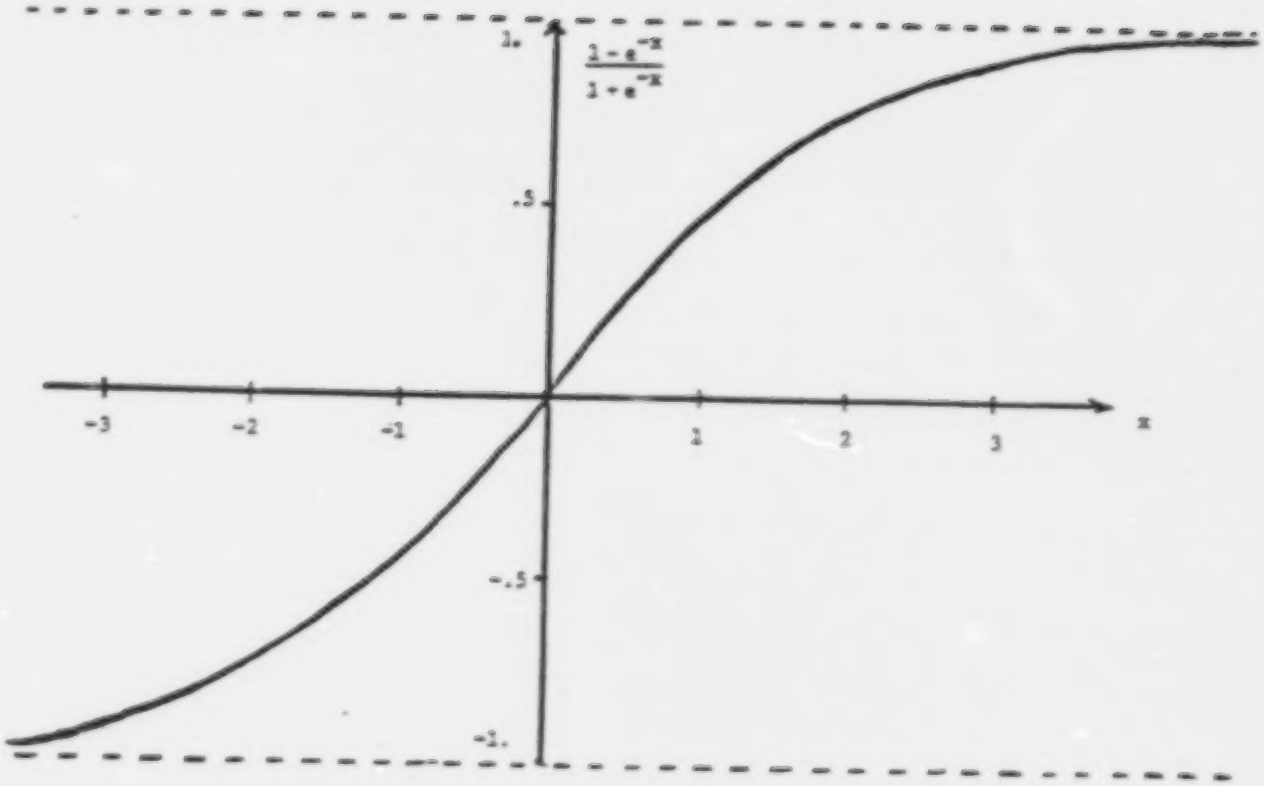


FIGURE 2



L'interprétation économique d'une telle formulation est bien fondée[1]. En effet, soit $\alpha_1 z_{ij}^1 = \alpha_1 d_{ij}^{-1}$. On s'attend à ce que la valeur absolue de la corrélation entre deux erreurs ε_i et ε_j soit une fonction inverse de la distance. Dans ce cas particulier, on a alors :

$$(53) \qquad r_{ij} = \frac{1 - e^{-\alpha_1/d_{ij}}}{1 + e^{-\alpha_1/d_{ij}}} \quad .$$

Alors, lorsque $d_{ij} \to \infty$ $\alpha_1/d_{ij} \to 0$ et donc $r_{ij} \to 0$. Lorsque la distance tend vers zéro, $\alpha_1/d_{ij} \to \infty$. Alors pour $\alpha_1 > 0$ $r_{ij} \to +1$ et pour $\alpha_1 < 0$ $r_{ij} \to -1$.

Afin de conserver la propriété de corrélation nulle entre les erreurs de zones très éloignées, la forme fonctionnelle que l'on propose s'écrit :

$$(54) \qquad r_{ij} = \frac{1 - e^{-d_{ij}^{-\lambda}(\alpha_1 + \alpha_2 v_{ij}^2 + \ldots + \alpha_H v_{ij}^H)}}{1 + e^{-d_{ij}^{-\lambda}(\alpha_1 + \alpha_2 v_{ij}^2 + \ldots + \alpha_H v_{ij}^H)}} \quad , \qquad \text{pour } i > j$$

où $v_{ij}^1 = 1 \ \forall i > j$, et où on pose $\theta = (\alpha_1, \alpha_2, \ldots, \alpha_H, \lambda)'$. Si on écrit $z_{ij}^h = d_{ij}^{-\lambda} v_{ij}^h$, $h = 1, \ldots, H$, on retrouve l'expression (52). La partie linéaire $\alpha_1 + \alpha_2 v_{ij}^2 + \ldots + \alpha_H v_{ij}^H$ permet à l'utilisateur d'introduire différentes variables afin d'expliquer avec l'aide d'une seule relation le lien entre les paires d'erreurs ε_i et ε_j de l'étude pour $i \neq j$.

---

1. Il est évident que toute forme fonctionnelle monotone f(x) telle que $f(-\infty) = -1$ $f(0) = 0$ et $f(\infty) = 1$ peut être pertinente.

4.2.2.1) *Quelques variables pour expliquer la corrélation spatiale*

Il est difficile ici de donner ou d'énumérer toutes les variables qu'il est intéressant d'introduire dans la forme linéaire de (54). Toutefois, c'est généralement le caractère économico-géographique spécifique de l'étude et l'intuition économique du chercheur qui lui dicteront quelles variables introduire dans le vecteur $V_{ij} = (1, V_{ij}^2, \ldots, V_{ij}^H)'$. Abstraction faite du signe de la corrélation r_{ij}, en se basant sur les principes de modélisation dans les modèles de flux, l'importance du lien, comme on l'a dit, risque fortement d'être une fonction inverse de la distance. Si on s'inspire de ce même type de modélisation, d'autres variables nous viennent à l'esprit.

i) Des variables à doubles indices

L'importance du lien entre deux zones i et j est généralement proportionnelle à certains des facteurs suivants :

1. le flux de passagers par autobus, auto ou par tout autre moyen de transport,
2. le degré d'utilisation des lignes téléphoniques entre les deux zones,
3. l'inverse du coût de transport mesuré, ce qui généralement se représente par une fonction qui tient compte du temps de trajet, du temps d'attente, etc.

ii) Autres variables

Dans les modèles d'explication de flux, on introduit souvent le produit $Pop_i \cdot Pop_j$ des populations en i et en j, le produit $Rev_i \cdot Rev_j$ des revenus en i et en j, ou tout autre produit de variables socio-économiques qui tient compte de l'importance de l'activité économique dans les deux zones. Plus l'activité est forte, plus le lien risque d'être fort. De la même façon, on pourrait prendre $Pop_i + Pop_j$ ou $Rev_i + Rev_j$.

Par l'introduction de variables dichotomiques (0, 1), on peut chercher à modéliser du même coup deux types de relation, c'est-à-dire soit $Z'_{ij}\alpha = d_{ij}^{-\lambda} (\alpha_1 + \alpha_2 p_{ij})$ où, par exemple, $p_{ij} = 1$, si $Pop_i > \overline{Pop}$ et $Pop_j > \overline{Pop}$ et $p_{ij} = 0$ ailleurs, et où $\overline{Pop}$ désigne la population moyenne de la région étudiée. De cette façon, on peut discriminer entre les zones à petite et à grande densité de population. Pour $\alpha_2 \neq 0$, on donne un niveau plus élevé à la corrélation qui concerne les zones à grande densité de population alors que les zones à faible population se modélisent par l'unique relation $\alpha_1 d_{ij}^{-\lambda}$. C'est ce qui explique pourquoi on soutient que c'est au chercheur que revient la tâche de décider quel genre de modélisation de la corrélation spatiale il doit faire.

De la même façon, grâce au principe de contiguïté populaire dans les recherches spatiales, on peut mettre $p_{ij} = 1$, si i et j sont contigües d'ordre 1 et $p_{ij} = 0$ si non. Alors de cette façon, la corrélation spatiale entre deux zones s'explique de façon non linéaire par une partie $\alpha_2 d_{ij}^{-\lambda}$ due à la contiguïté et un facteur $\alpha_1 d_{ij}^{-\lambda}$ commun à toutes les paires de zones. Un test de signification sur α_2 permet de tester si la contiguïté est un facteur significatif pour expliquer r_{ij}.

Pour ce qui concerne les signes attendus des coefficients, mis à part λ pour lequel on s'attend à un signe positif, il semble à priori impossible de donner une justification valable au signe des coefficients α_h $h = 1, 2, \ldots, H$. Cet argument est acceptable si on considère que dans l'estimation MCG du modèle qui intéresse le chercheur, ce qui importe, c'est de tenir compte de l'importance du lien entre les erreurs du modèle et non de spécifier ou de justifier le signe de la corrélation pour chacune des paires d'erreurs de l'étude.

Remarque

Pour la forme simplifiée $\alpha_1 d_{ij}^{-\lambda}$, si d'après la configuration spatiale des erreurs de l'estimation MCO, il semble y avoir corrélation positive (les résidus de même signe se regroupent), il paraît alors raisonnable de croire à priori à un signe positif pour α_1. Dans le cas d'éparpillement des valeurs similaires, on s'attendrait à un signe négatif.

On aura l'occasion de revenir sur l'interprétation économique de la forme (54) dans le chapitre consacré à l'étude des résultats d'estimation du modèle que l'on désire estimer.

4.2.3) *Une procédure du maximum de vraisemblance*

Comme on l'a déjà souligné, lorsque la loi de distribution conjointe de l'échantillon est normale, l'estimation du maximum de vraisemblance produit généralement une estimation $\tilde{\beta}$ de type MCGE pour le vecteur β de l'équation (10)

$$\underset{(N\times 1)}{y} = \underset{(N\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(N\times 1)}{\varepsilon}$$

où $\varepsilon \sim N(0, \sigma^2 R)$ et où d'après l'équation (54), $r_{ij} = \dfrac{1 - e^{-Z'_{ij}\alpha}}{1 + e^{-Z'_{ij}\alpha}}$, où

$Z'_{ij}\alpha = d_{ij}^{-\lambda}(\alpha_1 + \alpha_2 V_{ij}^2 + \ldots + \alpha_H V_{ij}^H)$, et où on pose :

$\theta = (\alpha_1, \alpha_2, \ldots, \alpha_H, \lambda)$.

L'algorithme de solution pour ce problème est décrit au prochain chapitre. Avant de procéder à l'élaboration de l'algorithme, arrêtons-nous sur certains points :

1. Sous certaines conditions générales, le vecteur $\tilde{\xi} = (\tilde{\beta}', \tilde{\theta}', \tilde{\sigma}^2)$ est convergent, asymptotiquement, normalement distribué et efficace. Pour ces conditions, voir Magnus (1978), section 6) ou section 5) de ce chapitre.

2. Comme la matrice R doit être définie positive pour toute valeur du vecteur $\tilde{\theta}$, on a interdit à l'algorithme de considérer comme une solution possible tout vecteur estimé de θ qui ne respecte pas cette condition. On le verra, ceci va toutefois nous causer de nombreux problèmes car ce n'est pas la façon la plus orthodoxe d'introduire ce genre de contraintes[1].

---

[1]. Pour une variable scalaire x, la méthode la plus fiable pour introduire une contrainte de non négativité consiste à faire travailler l'algorithme dans l'espace de x^2. Pour plus d'information sur le genre de modification à apporter, consulter la section 5) du chapitre prochain. Dans le cas qui nous concerne ici, on a été incapable d'imaginer une astuce satisfaisante.

3. Un habitué du calcul numérique sur ordinateur comprend rapidement que la présence des exponentielles à l'équation (54) risque de causer des problèmes numériques. En effet, le domaine de $Z'_{ij}\alpha$ est alors assez restreint. Sur C.D.C., on doit respecter $-741{,}67 \leq Z'_{ij}\alpha \leq 675{,}84$, pour que puisse s'évaluer $e^{-Z'_{ij}\alpha}$. Afin de contourner ce problème, on suggère de remplacer l'expression (54) par une fonction dont le domaine va de $-\infty$ à $+\infty$ et qui respecte en même temps le fait que $-1 \leq r_{ij} \leq 1$. Soit :

$$r_{ij} = \tanh (Z'_{ij}\alpha) \quad . \tag{55}$$

Cette relation est très intéressante puisqu'il existe une relation directe entre les expressions (54) et (55). En effet (C.R.C. Standard Mathematical Tables, 26ième édition, p. 214), on a :

$$\tanh (x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \equiv \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad . \tag{56}$$

En remplacant x par $Z'_{ij}\alpha$, on obtient :

$$\tanh (Z'_{ij}\alpha) = \frac{1 - e^{-2Z'_{ij}\alpha}}{1 + e^{-2Z'_{ij}\alpha}} \quad . \tag{57}$$

On peut donc conclure que les relations (54) et (57) donneront exactement la même estimation $\tilde{r}_{ij}$ pour r_{ij}, sauf que dans la deuxième expression, les estimations de α seront deux fois plus petites qu'en (54). Dû au fait qu'avec $\tanh (Z'_{ij}\alpha)$, $Z'_{ij}\alpha$ peut être compris entre $-\infty$ et $+\infty$, on modélisera la corrélation spatiale par la relation suivante :

(58) $\quad r_{ij} = \tanh\,(Z'_{ij}\alpha)$,

où $Z'_{ij}\alpha = d_{ij}^{-\lambda}\,(\alpha_1 + \alpha_2\, v_{ij}^2 + \ldots + \alpha_H\, v_{ij}^H)$ avec $\theta = (\alpha_1, \alpha_2, \ldots, \alpha_H, \lambda)'$.

Avant de passer à la prochaine section, rappelons certaines relations fondamentales rattachées à la tangente hyperbolique.

Soit $-1 < y < 1$ et $-\infty < x < \infty$, on a : $y = \tanh\,(x) = \dfrac{1 - e^{-2x}}{1 + e^{-2x}}$

dont la relation inverse s'écrit : $x = \tanh^{-1}\,(y) = \frac{1}{2}\ln\,(\frac{1+y}{1-y})$ pour $y^2 < 1$. On a aussi : $\dfrac{\partial \tanh\,(x)}{\partial u} = \operatorname{sech}^2\,(x) \cdot \dfrac{\partial x}{\partial u}$, où $\operatorname{sech}^2\,(x) = 1 - \tanh^2\,(x)$.

5) <u>Sur les propriétés asymptotiques des estimateurs du maximum de vraisemblance pour le modèle MCG spatial</u>

Grâce aux résultats de Magnus (1978), la méthode du maximum de vraisemblance proposée ici produira des estimateurs convergents, asymptotiquement normaux et efficaces si les hypothèses H1 à H7 de la section 7.1) du chapitre III sont respectées.

Comme hypothèse critique, l'hypothèse H2 stipule que la matrice $\sigma^2\, R(\theta)$, $\theta = (\alpha_1, \ldots, \alpha_H\ \lambda)'$ doit être définie positive pour toute valeur de θ, $\theta \in \Theta$. Bien que l'algorithme que l'on propose rejette tout vecteur tel que $R(\tilde{\theta})$ est non définie positive, le fait qu'il existe de telles valeurs de θ montre que l'hypothèse H2 est fort probablement violée. Nos estimations du maximum de vraisemblance risquent donc d'être non convergentes. A faute de mieux, examinons tout de même, au chapitre XII, les résultats d'estimation de notre procédure. Mais avant, décrivons en détail, l'algorithme.

CHAPITRE XI

MÉTHODE DES MOINDRES CARRÉS GÉNÉRALISÉS (MCG) DU MAXIMUM DE VRAISEMBLANCE COMME SOLUTION À UN PROBLÈME DE CORRÉLATION SPATIALE

Un modèle linéaire dont les erreurs sont spatialement autocorrelées s'écrit :

$$(1) \qquad \underset{(N\times 1)}{y} = \underset{(N\times K)}{X} \; \underset{(K\times 1)}{\beta} + \underset{(N\times 1)}{\varepsilon}$$

où $\varepsilon \sim N(0, \sigma^2 R)$.

Comme forme fonctionnelle pour modéliser la corrélation spatiale, on postule :

$$(2) \qquad r_{ij} \begin{cases} = \tanh\,(Z'_{ij}(\lambda) \cdot \alpha) = \dfrac{1 - e^{-2Z'_{ij}(\lambda)\,\cdot\,\alpha}}{1 + e^{-2Z'_{ij}(\lambda)\,\cdot\,\alpha}} \quad , & \text{si } i > j \ , \\ = 1 \quad , & \text{si } i = j \ , \end{cases}$$

et où finalement :

$$(3) \qquad Z'_{ij}(\lambda) \cdot \alpha = d_{ij}^{-\lambda}\,(\alpha_1 + \alpha_2\, V_{ij}^2 + \ldots + \alpha_H\, V_{ij}^H) = d_{ij}^{-\lambda}\, V'_{ij}\, \alpha \quad ,$$

où V_{ij} est un vecteur colonne de H éléments, et V_{ij}^h la hième variable explicative de la forme linéaire formant (3). Pour simplifier la notation, posons $Z_{i,j}(\lambda) = Z_{i,j}$ et appelons chaque indice i,j (i > j) l'indice t. On a donc à l'équation (3) : $Z'_t\,\alpha = d_t^{-\lambda}\, V'_t\,\alpha$, t=1,...,T, où $T = \frac{N\times(N-1)}{2}$. Soit θ le vecteur des paramètres qui forme la matrice de corrélation spatiale R :

$$(4) \qquad \underset{([H+1]\times 1)}{\theta} = (\alpha_1, \alpha_2, \ldots \alpha_H, \lambda)' \quad ,$$

on a :

---

[1]. A cause de la symétrie de la matrice de corrélation R, on n'estimera que les $\frac{N\times(N-1)}{2}$ éléments sous la diagonale.

1) Fonction de vraisemblance et dérivées premières

$$(5) \qquad \mathcal{L}(\beta, \sigma^2, \theta|y) = (2\pi \cdot \sigma^2)^{-N/2} |R|^{-1/2} \exp\{-\frac{(y - X\beta)' R^{-1} (y - X\beta)}{2\sigma^2}\}$$

dont le log naturel s'écrit :

$$(6) \qquad L(\beta, \sigma^2, \theta|y) = -\frac{N}{2} \ln (2\pi) - \frac{N}{2} \ln \sigma^2 - \frac{1}{2} \ln |R| - \frac{(y - X\beta)' R^{-1} (y - X\beta)}{2\sigma^2}$$

où $\varepsilon = (y - X\beta)$. D'après les conditions du premier ordre, on a :

$$(7) \qquad \frac{\partial L}{\partial \beta} = \frac{X' R^{-1} (y - X\beta)}{\sigma^2} \quad ,$$

$$(8) \qquad \frac{\partial L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{(y - X\beta)' R^{-1} (y - X\beta)}{2\sigma^4} \quad ,$$

$$(9) \qquad \frac{\partial L}{\partial \theta_i} = -\frac{1}{2} \text{tr} (R^{-1} \frac{\partial R}{\partial \theta_i}) + \frac{1}{2\sigma^2} \varepsilon' R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1} \varepsilon \quad , \qquad i = 1, \ldots, H+1.$$

Egalant à zéro les équations (7) et (8), on obtient :

$$(10) \qquad \beta^* = (X' R^{-1} X)^{-1} X' R^{-1} y \quad ,$$

et

$$(11) \qquad \sigma^{*2} = \frac{(y - X\beta^*)' R^{-1} (y - X\beta^*)}{N} \quad .$$

β^* et σ^{*2} d'après (10) et (11) dépendent exclusivement du vecteur θ. Pour toute valeur de θ, les équations (10) et (11) rendent les dérivées (7) et (8) nulles. Ceci nous permet alors de concentrer la fonction (6) par rapport à β et σ^2. On obtient :

(12) $$L^{**}(\theta|\beta^*, \sigma^{*2}, y) = -\frac{N}{2}[\ln(2\pi) + 1] - \frac{N}{2}\ln\sigma^{*2} - \frac{1}{2}\ln|R| \quad ,$$

laquelle doit être maximisée par rapport au vecteur $\theta = (\alpha_1, \alpha_2, \ldots \alpha_H, \lambda)'$ seulement. Il faut donc trouver les valeurs de θ_i qui, à la fois, maximisent (12) et rendent l'expression suivante nulle :

(13) $$\frac{\partial L^{**}}{\partial \theta_i} = -\frac{1}{2}\,\mathrm{tr}\left(R^{-1}\frac{\partial R}{\partial \theta_i}\right) + \frac{(y - X\beta^*)'}{2\sigma^{*2}} R^{-1}\frac{\partial R}{\partial \theta_i} R^{-1}(y - X\beta^*) \quad , \qquad i=1,\ldots,H+1$$

Lorsque l'expression (13) est nulle $\forall\theta_i$ évalué en $\hat{\theta}_i$, les expressions β^* et σ^{*2} évaluées en $\hat{\theta}$ produisent les estimations du maximum de vraisemblance de β et σ^2. On a alors $\hat{\beta} = (X'\,\hat{R}^{-1}\,X)^{-1}\,X'\,\hat{R}^{-1}\,y$ et $\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'\,\hat{R}^{-1}\,(y - X\hat{\beta})}{N}$. Afin d'évaluer l'équation (13), on doit connaître $\frac{\partial R}{\partial \theta_i}$. D'après nos conventions, les éléments sous la diagonale de la matrice des corrélations spatiales s'écrivent :

(14) $$r_t = \tanh(Z_t'\,\alpha) \quad ,$$

où $Z_t'\,\alpha = d_t^{-\lambda}(V_t'\,\alpha)$. On a donc :

$$\frac{\partial r_t}{\partial \theta_i} = \frac{\partial \tanh(Z_t'\,\alpha)}{\partial Z_t'\,\alpha} \cdot \frac{\partial Z_t'\,\alpha}{\partial \theta_i} = \mathrm{sech}^2(Z_t'\,\alpha) \cdot \frac{\partial Z_t'\,\alpha}{\partial \theta_i} = [1 - \tanh^2(Z_t'\,\alpha)]\frac{\partial Z_t'\,\alpha}{\partial \theta_i} \quad ,$$

d'où :

(15) $$\frac{\partial r_t}{\partial \theta_i} = (1 - r_t^2) \cdot \frac{\partial Z_t'\,\alpha}{\partial \theta_i} \quad ,$$

avec

(15.1) $\theta_i =$

	$\frac{\partial Z_t' \alpha}{\partial \theta_i}$
α_h	$d^{-\lambda} V_t^h = Z_t^h$
λ	$- V_t' \alpha \cdot d^{-\lambda} \ln (d_t) = - Z_t' \alpha \cdot \ln (d_t)$

,

où h = 1,...,H .

2) Dérivées secondes

De (7), (8) et (9), on obtient :

(16) $$\frac{\partial^2 L}{\partial \beta \partial \beta'} = - \frac{X' R^{-1} X}{\sigma^2} \quad ,$$

(17) $$\frac{\partial^2 L}{\partial \beta \partial \sigma^2} = - \frac{X' R^{-1} \varepsilon}{\sigma^4} \quad ,$$

(17.1) $$\frac{\partial^2 L}{\partial \beta \partial \theta_i} = - \frac{X'}{\sigma^2} (R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1}) \varepsilon \quad ,$$

(18) $$\frac{\partial^2 L}{\partial \sigma^2 \partial \sigma^2} = - \frac{N}{2\sigma^4} - \frac{\varepsilon' R^{-1} \varepsilon}{\sigma^6} \quad ,$$

(19) $$\frac{\partial^2 L}{\partial \sigma^2 \partial \theta_i} = - \frac{\varepsilon'}{2\sigma^4} (R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1}) \varepsilon \quad ,$$

$$(20)\quad \frac{\partial^2 L}{\partial\theta_i \partial\theta_j} = -\frac{1}{2}\,\mathrm{tr}\left(-R^{-1}\frac{\partial R}{\partial\theta_j}R^{-1}\frac{\partial R}{\partial\theta_i} + R^{-1}\frac{\partial^2 R}{\partial\theta_i\partial\theta_j}\right)$$

$$-\frac{\varepsilon'}{\sigma^2}R^{-1}\left(\frac{\partial R}{\partial\theta_j}R^{-1}\frac{\partial R}{\partial\theta_i}\right)R^{-1}\varepsilon + \frac{1}{2\sigma^2}\varepsilon' R^{-1}\frac{\partial^2 R}{\partial\theta_i\partial\theta_j}R^{-1}\varepsilon \quad .$$

Dans l'évaluation de l'équation (20), on a besoin de $\frac{\partial^2 R}{\partial\theta_i\partial\theta_j}$. Si on se rappelle les équations (15) et (15.1) où

$\frac{\partial r_t}{\partial\theta_i} = (1 - r_t^2)\frac{\partial Z_t'\alpha}{\partial\theta_i}$, $i = 1,\ldots,H+1$, où $t = 1,\ldots,\frac{N\times(N-1)}{2}$, on a alors :

$$(21)\quad \frac{\partial^2 r_t}{\partial\theta_i\partial\theta_j} = -2\,r_t\frac{\partial r_t}{\partial\theta_j}\cdot\frac{\partial Z_t'\alpha}{\partial\theta_i} + (1 - r_t^2)\frac{\partial^2 Z_t'\alpha}{\partial\theta_i\partial\theta_j} \quad ,$$

où, grâce à (15.1), avec $h = 1,\ldots,H$ et $g = 1,\ldots,H$,

(21.1) $\theta_i =$

$\frac{\partial^2 Z_t'\alpha}{\partial\theta_i\partial\theta_j}$	$\theta_j =$ α_h	λ
α_g	0	$-Z_t^g \ln(d_t)$
λ	$-Z_t^h \ln(d_t)$	$Z_t'\alpha \ln^2(d_t)$

3) Simplification des formules pour la programmation

En nous servant de (15), (15.1), (21) et (21.1), on a :

$$(22) \qquad \frac{\partial^2 r_t}{\partial \alpha_h \partial \alpha_g} = -2\, r_t\, (1-r_t^2)\, Z_t^g\, Z_t^h \quad .$$

$$(23) \qquad \frac{\partial^2 r_t}{\partial \alpha_h \partial \lambda} = 2\, r_t\, (1-r_t^2)\, Z_t'\alpha \ln (d_t)\, Z_t^h - (1-r_t^2)\, Z_t^h \ln (d_t)$$

$$= (1-r_t^2)\, Z_t^h \ln (d_t)\, [2\, r_t \cdot Z_t'\alpha - 1] \quad .$$

$$(24) \qquad \frac{\partial^2 r_t}{\partial \lambda \partial \lambda} = -2\, r_t\, (1-r_t^2)\, (Z_t'\alpha)^2 \ln^2 (d_t) + (1-r_t^2)\, Z_t'\alpha \ln^2 (d_t)$$

$$= (1-r_t^2)\, Z_t'\alpha \ln^2 (d_t)\, [1 - 2\, r_t\, Z_t'\alpha] \quad .$$

en remarquant que $Z_t'\alpha = \tanh^{-1} (r_t)$.

4) Calcul de la matrice variance-covariance des paramètres estimés

Comme par les équations (17) et (17.1), $E(\frac{\partial^2 L}{\partial \beta \partial \sigma^2}) = 0$ et $E(\frac{\partial^2 L}{\partial \beta \partial \theta_i}) = 0$, $i = 1,\ldots,H+1$, on vérifie que $\hat{\beta}$ est asymptotiquement indépendant de $\hat{\sigma}^2$ et $\hat{\theta}$. Grâce à la partition de la matrice d'information, la matrice variance-covariance de $\hat{\beta}$ se calcule par la négative de l'inverse numérique de la matrice de l'équation (16). Pour $\hat{\sigma}^2$ et $\hat{\theta}$, on se limite à la sous-matrice formée par les équations (18) à (20). Ainsi on trouve :

$$V(\hat{\beta}) = \hat{\sigma}^2 \, (X' \, \hat{R}^{-1} \, X)^{-1}$$

et

$$V\begin{pmatrix}\hat{\sigma}^2 \\ \hat{\theta}\end{pmatrix} = \left[- \frac{\partial^2 L}{\partial \begin{bmatrix}\sigma^2 \\ \theta\end{bmatrix} \partial \, [\sigma^2 \; \theta']} \right]^{-1} \Bigg|_{\substack{\theta = \hat{\theta} \\ \sigma^2 = \hat{\sigma}^2}} .$$

L'évaluation de l'espérance mathématique des équations (16) à (20) étant simple, l'algorithme d'estimation calcule aussi :

$$V\begin{pmatrix}\hat{\sigma}^2 \\ \hat{\theta}\end{pmatrix} = \left[-E \left(\frac{\partial^2 L}{\partial \begin{bmatrix}\sigma^2 \\ \theta\end{bmatrix} \partial \, [\sigma^2 \; \theta']} \right) \right]^{-1} \Bigg|_{\substack{\theta = \hat{\theta} \\ \sigma^2 = \hat{\sigma}^2}} .$$

En effet :

$$(25) \qquad - E \left(\frac{\partial^2 L}{\partial \beta \partial \beta'} \right) = \frac{X' \, R^{-1} \, X}{\sigma^2} \quad ,$$

$$(26) \qquad - E \left(\frac{\partial^2 L}{\partial \sigma^2 \partial \sigma^2} \right) = \frac{N}{2\sigma^4} \quad ,$$

$$(27) \qquad - E \left(\frac{\partial^2 L}{\partial \sigma^2 \partial \theta_i} \right) = \frac{1}{2\sigma^2} \, tr \left(R^{-1} \frac{\partial R}{\partial \theta_i} \right) \quad ,$$

$$(28) \qquad - E \left(\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right) = \frac{1}{2} \, tr \left(R^{-1} \frac{\partial R}{\partial \theta_j} \; R^{-1} \frac{\partial R}{\partial \theta_i} \right) \quad ,$$

car $E(\varepsilon\varepsilon') = \sigma^2 R$ et où :

$$(29) \qquad -E\left(\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\right) = -E\left(\frac{\partial^2 L}{\partial\beta\partial\theta_i}\right) = \underset{(K\times 1)}{0} \quad .$$

5) Changement d'espace pour les paramètres

Lorsqu'on veut introduire à priori des contraintes d'inégalité stricte de type :

$$(30) \qquad -h < x < h$$

dans un algorithme, on peut le faire de feux façons. La méthode la plus directe, mais la moins recommandée, suggère d'introduire dans le programme une ligne supplémentaire qui l'empêche de choisir une valeur de x en dehors du domaine défini par (30). Comme solution plus acceptable, il existe le changement d'espace.

Soit z une variable distribuée sur le domaine $-\infty$ à $+\infty$. Comme :

$$(31) \qquad -1 < \tanh(z) < 1 \quad ,$$

la variable $x = h \cdot \tanh(z)$ sera comprise entre $-h$ et h.

Alors bien que x soit limité au domaine $-h < x < h$, l'algorithme travaillera sur le domaine non contraint $-\infty < z < \infty$. Pour ce faire, il faut fournir à l'algorithme les dérivées premières en terme de la variable non contrainte z. On a donc les relations suivantes :

(32) $\quad x = h \tanh(z) \quad \rightarrow z = \tanh^{-1}(x/h)$

$$z = \frac{1}{2} \ln \left(\frac{1 + x/h}{1 - x/h}\right)$$

avec

(33) $\quad \frac{\partial L}{\partial z} = \frac{\partial L}{\partial x} \cdot \frac{\partial x}{\partial z}$, [$x = f(z)$ étant une relation monotone]

où

(34) $\quad \frac{\partial x}{\partial z} = h \frac{\partial \tanh(z)}{\partial z} = h \operatorname{sech}^2(z) = h (1 - \tanh^2(z)) = h \left[1 - \left(\frac{x}{h}\right)^2\right]$.

Comme application généralement utilisée, soit $|\rho| < 1$, en posant $\rho = \tanh(z)$ où $-\infty < z < \infty$, on a, grâce à (32) et (34) pour $x = \rho$ et $h = 1$: $\frac{\partial \rho}{\partial z} = (1 - \rho^2)$, où $z = \frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho}\right)$. C'est d'ailleurs cette dernière transformation que l'on utilise pour estimer le ρ du modèle SAR(1).

CHAPITRE XII

RÉSULTATS D'ESTIMATION DE LA PROCÉDURE DES MOINDRES CARRÉS GÉNÉRALISÉS DU MAXIMUM DE VRAISEMBLANCE

Au cours de ce chapitre, nous porterons notre attention sur différentes estimations des moindres carrés généralisés (MCG) du modèle linéaire traité aux chapitres V à VII. Comme modélisation de la corrélation spatiale, on s'attardera premièrement à la forme fonctionnelle suivante :

$$(1) \qquad r_{ij} = \tanh\left(\alpha_1 \, d_{ij}^{-\lambda}\right) \quad , \qquad i > j$$

selon laquelle la corrélation entre l'erreur ε_i et l'erreur ε_j est une fonction inverse de la distance. L'algorithme trouvera alors une solution intéressante en quelques itérations. C'est lorsqu'on ajoutera des variables supplémentaires à l'équation (1) que le problème s'envenimera. En effet, pour certaines spécifications, l'algorithme sera incapable de converger.

1) Résultats d'estimation de la forme $r_{ij} = \tanh\left(\alpha_1 \, d_{ij}^{-\lambda}\right)$

$$\underset{(1,072)}{y = 7,325} + \underset{(2,4363)}{0,00206\ X}$$

$$(2) \qquad \hat{\alpha}_1 = \underset{(6,7496)}{2,4725}$$

$$\hat{\lambda} = \underset{(2,5337)}{0,5612}$$

$R^2 = 0,2717 \qquad \hat{\sigma}^2 = 44,4203 \qquad \ln \mathcal{L} = - 24,429 \qquad |R| = 2,40844 \ E\text{-}22$.

Selon l'évidence statistique du test t, la puissance de la distance est égale à - 0,5612, ce qui répond à nos attentes. L'estimation MCG du modèle conduit pour α et β à des résultats relativement comparables à ceux obtenus par une méthode SAR(1) avec matrice de contiguïté booléenne normalisée symétrique. C'est ce qu'on présente au tableau 1.

TABLEAU 1

	MCG maximum de vraisemblance	SAR(1) $\underset{\sim}{W}$ booléenne	MCO
α	7,32519 (1,07596)	5,28736 (1,27515)	-8,4478 (-2,7538)
β	0,00205 (2,436347)	0,00260 (4,3991)	0,00527 (7,7263)

Dans ces estimations, l'ordonné à l'origine est non significative alors que la pente donne une valeur raisonnablement près de celle obtenue par l'estimation du SAR(1) résiduel. On limite à α et β la comparaison entre les résultats d'estimations. En effet, des différences fondamentales dans la formulation des modèles respectifs rendent toute autre comparaison plus ardue. Par exemple, dans le présent modèle, $\hat{\sigma}^2 = 44,4203$ représente l'estimation de la variance homoscédastique $\sigma^2_{y_i} = \sigma^2_{\varepsilon_i} = \sigma^2$, pour $i = 1,\ldots,N$. Pour le modèle SAR(1) qui, on le rappelle, s'écrit $y = X\beta + \varepsilon$ où $\varepsilon = \rho W \varepsilon + u$, σ^2 représente plutôt la variance homoscédastique de u_i. Pour $\sigma^2_{y_i} = \sigma^2_{\varepsilon_i}$, il faut prendre l'élément ii de la matrice $\sigma^2 V = \sigma^2 [(I - \rho W)' (I - \rho W)]^{-1} = \sigma^2 [I - 2\rho W + \rho^2 W^2]^{-1}$. Dans ce cas, on n'a sûrement pas homoscédasticité. Afin de pouvoir comparer avec le présent modèle MCG, on pourrait calculer la variance estimée moyenne suivante : $\overline{\sigma^2_y} = \hat{\sigma}^2 \cdot \frac{\sum_{i=1}^{N} v_{ii}}{N}$ où v_{ii} représente l'élément ii de la matrice V.

Pour des valeurs de départ $\alpha_1^o = 0,0$ et $\lambda^o = 2,0$, l'algorithme a trouvé la solution en 17 itérations, ce qui a nécessité toutefois 60 évaluations de la fonction. Sur ces 60 appels, seulement 8 ont produit un vecteur de paramètres non permis. Ces derniers appels se sont d'ailleurs produits au cours des 15 premières évaluations de la fonction. Le même vecteur de solution $\hat{\theta} = [\hat{\alpha}_1 \ \hat{\lambda}]' = [2,4725 \ \ 0,5612]'$ a été retrouvé pour plusieurs autres valeurs de départ.

Afin de s'assurer de la crédibilité de nos résultats, on présente, aux figures 1 et 2 suivantes, des grilles sur des régions de α_1 et λ raisonnables.

FIGURE 1

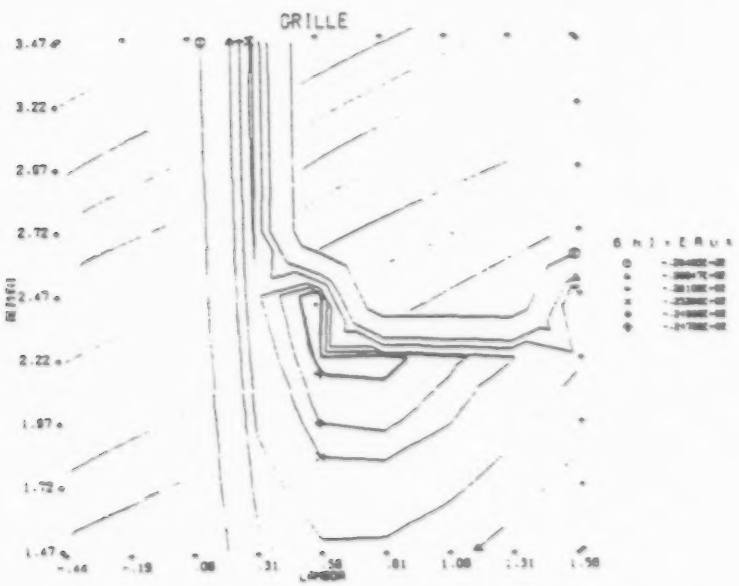
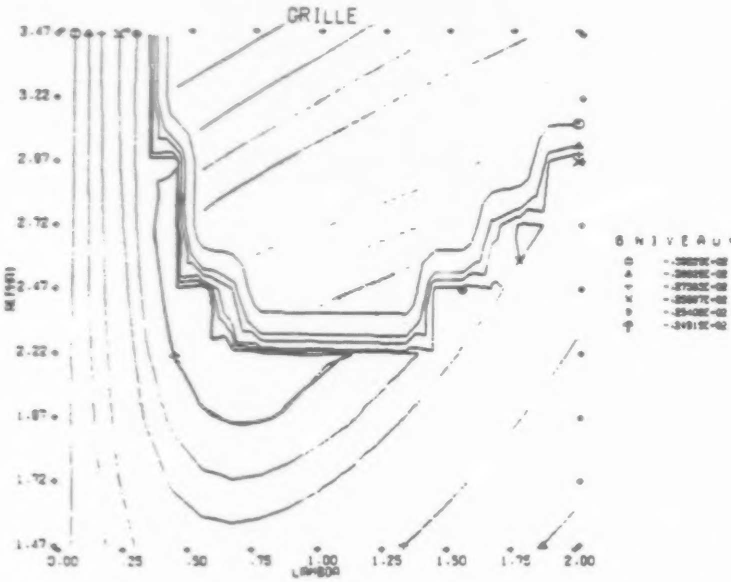


FIGURE 2

FMAXI,FMINI -24.4301216886 -37.62772171613

1	3.4723	-37.6277	-28.1835	-23.8347	-23.0720	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000
2	3.2223	-37.3774	-28.2927	-23.8631	-24.9848	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000
3	2.9723	-37.1270	-28.4331	-23.9469	-24.9821	-24.9323	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000
4	2.7223	-36.8762	-28.6139	-26.0918	-23.0498	-24.3490	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000
5	2.4723	-36.6249	-28.8340	-26.3039	-23.2047	-24.6729	-24.4301	-27.3990	-38.0000	-38.0000	-38.0000
6	2.2223	-36.3737	-29.1016	-26.5998	-23.4617	-24.9030	-24.6247	-24.5176	-24.3643	-24.7173	-24.8393
7	1.9723	-36.1192	-29.4200	-26.9839	-23.8273	-23.3643	-24.9873	-24.8803	-24.8843	-24.9699	-23.1218
8	1.7223	-35.8632	-29.8094	-27.4786	-26.2491	-23.7800	-23.3138	-23.4280	-23.4633	-23.3846	-23.7729
9	1.4723	-35.6033	-30.3638	-28.0917	-27.0123	-26.4634	-26.2168	-26.1314	-26.2871	-26.3472	-36.5483
		0.0000	.1111	.2222	.3333	.4444	.5556	.6667	.7778	.8889	1.0000

1	3.4723	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000
2	3.2223	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000
3	2.9723	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-27.2062	-25.8794
4	2.7223	-38.0000	-38.0000	-38.0000	-38.0000	-38.0000	-28.0382	-25.6168	-25.9643	-26.4009
5	2.4723	-38.0000	-38.0000	-38.0000	-23.3243	-25.3993	-25.7211	-26.1209	-26.5372	-26.9308
6	2.2223	-24.9139	-23.0472	-23.2904	-23.3986	-25.9481	-26.3304	-26.7216	-27.1112	-27.4917
7	1.9723	-25.3327	-25.3961	-25.9025	-26.2901	-26.3968	-26.9620	-27.2878	-27.6860	-28.0326
8	1.7223	-26.0140	-26.2933	-26.6033	-26.9338	-27.2711	-27.6099	-27.9441	-28.2693	-28.5823
9	1.4723	-26.7930	-27.0680	-27.3622	-27.6667	-27.9743	-28.2794	-28.5777	-28.8639	-29.1430
		1.1111	1.2222	1.3333	1.4444	1.5556	1.6667	1.7778	1.8889	2.0000



Pour des valeurs de λ entre -0,44 et 1,56 et α_1 entre 1,47 et 3,47, on trouve comme seule portion intéressante celle qui est délimitée par les valeurs suivantes : $2,0 \leq \alpha_1 \leq 2,6$ et $0,4 \leq \lambda \leq 1,0$.

Les zones vides hachurées représentent les régions où les paires de valeur de $\hat{\alpha}_1$ et $\hat{\lambda}$ donnent une matrice R non positive définie[1]. A la figure 2, on présente la grille pour des valeurs de λ entre 0 et 2 et α_1 entre 1,47 et 3,47.

On n'obtient alors qu'une zone d'indétermination. On peut de plus remarquer, d'après le tableau de la figure 2 que la solution la plus haute (-24,4301) se situe à la frontière de la zone d'indétermination (-38,000) [2].

Les procédures MCG et SAR(1) dont on parle ici produisent des matrices de corrélations spatiales des résidus assez comparables. C'est ce que l'on présente au tableau 2 suivant.

---

[1]. A chaque évaluation de la fonction, on doit inverser la matrice de corrélation R. Pour ce faire, on utilise la sous-routine d'inversion de matrices définies positives LINV2P d'I.M.S.L. A chaque fois que la matrice R est algorithmiquement non définie positive, cette sous-routine renvoie un message d'erreur. C'est à ce moment qu'on rejète comme solution possible l'ensemble des valeurs qui forment au cours de cet appel, le vecteur θ.

[2]. Une fois l'évaluation de tous les points terminée, on fixe le niveau de la fonction de vraisemblance pour laquelle la matrice R est non définie positive à $\ln \mathcal{L}_{min} - 10$.

TABLEAU 2

A. MCG du maximum de vraisemblance

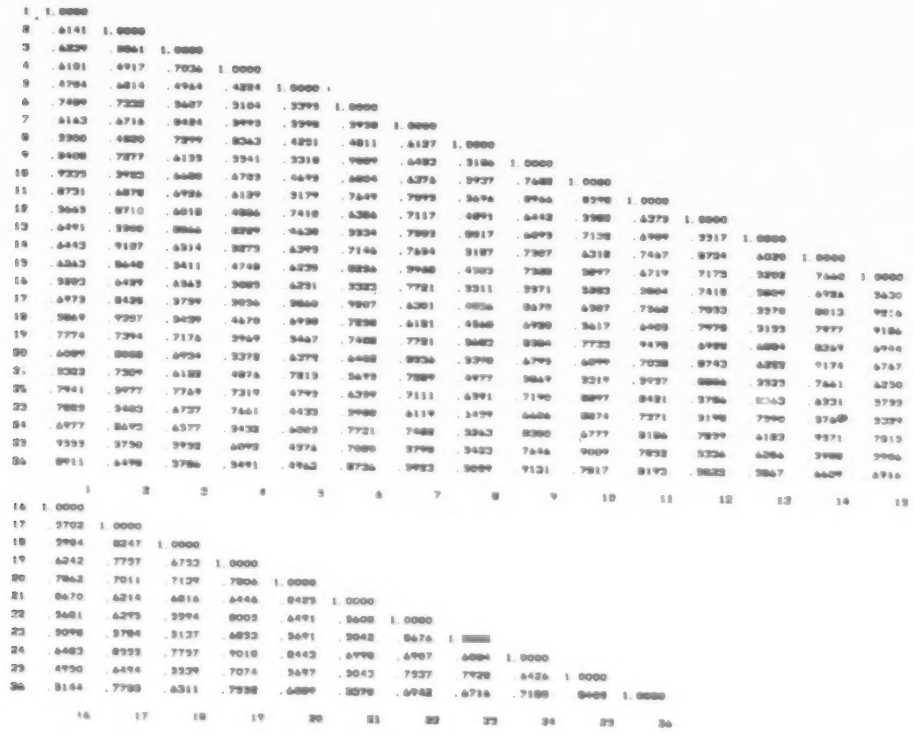
1	1.0000														
2	3033	1.0000													
3	3772	4443	1.0000												
4	6143	3946	8412	1.0000											
5	3166	6306	3238	2733	1.0000										
6	7996	6223	4792	4819	3729	1.0000									
7	6121	6484	7899	6823	3093	3770	1.0000								
8	3231	3616	8875	8088	2384	4152	6648	1.0000							
9	8636	6396	3634	3696	3973	7767	6615	6970	1.0000						
10	8761	4719	6483	7233	3087	6649	6384	6230	7710	1.0000					
11	8734	3611	6629	6783	3679	7306	7083	5974	8277	8733	1.0000				
12	4276	8324	4372	3723	8483	6990	6710	3491	3326	4173	4944	1.0000			
13	3669	3906	8993	9019	2778	4333	6996	7131	3379	6682	6446	3791	1.0000		
14	4639	8797	4352	3941	6925	5483	6817	3668	3814	4301	3238	8789	3948	1.0000	
15	4709	7453	3634	3353	4221	6147	5003	3008	6082	4236	5048	5583	3263	6123	1.0000
16	4628	6648	3361	4663	3848	4811	8211	4488	3378	4784	5425	7369	4769	7262	4706
17	6460	8731	4981	4648	3163	7947	6397	4163	8023	5886	6849	6816	4508	7493	8483
18	4846	8271	3846	3521	4824	6281	5406	3173	6257	4412	5243	6335	3441	6883	8946
19	7332	7527	6764	6346	3233	7399	8404	5803	8203	7269	8272	6922	4739	7384	6376
20	5230	8030	3640	4844	6786	3673	8246	4376	6238	5244	6069	8700	4894	8718	5747
21	4219	7131	4723	3961	6770	4646	7230	3767	3094	4243	4941	8341	4028	7774	4837
22	7433	8233	8667	8740	3668	6102	8137	8068	7160	8407	8313	4927	8403	5210	6438
23	7703	4340	7872	8326	2904	3907	6640	7110	6849	8778	7826	3929	7484	6210	3839
24	5343	8873	3086	4343	6131	6447	7213	4179	6828	5309	6252	8003	4503	8822	6862
25	8872	4319	3914	6336	2869	6979	3927	5331	7778	8888	8096	3883	5927	6220	4192
26	8823	5213	5193	5489	3190	8326	5726	4677	8784	7793	7937	4301	5033	6712	5013

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	1.0000														
17	6031	1.0000													
18	5214	8722	1.0000												
19	7198	8367	6747	1.0000											
20	8748	7873	6366	8182	1.0000										
21	8764	6120	5452	6839	8840	1.0000									
22	5916	6086	6671	7987	6247	5162	1.0000								
23	6681	5384	5992	6838	5018	6072	8679	1.0000							
24	7132	8444	7483	8223	8602	7231	5981	6931	1.0000						
25	6342	5806	6317	6792	6830	5889	7530	8462	5024	1.0000					
26	6664	6806	5124	7101	5136	6150	6705	6983	5613	8308	1.0000				

	16	17	18	19	20	21	22	23	24	25	26

TABLEAU 2 (suite)

B. SAR(1) avec matrice booléenne normalisée symétrique



Remarque

Les chiffres soulignés désignent les corrélations entre les paires d'erreurs contiguës. Soit S = diag [vâr (ε_1),...,vâr (ε_N)], la matrice $\hat{R}$ du SAR(1) est calculée comme suit :

(3) $$\hat{R} = \hat{\sigma}^2 \cdot S^{-1/2} (I - \hat{\rho} W)^{-1} (I - \hat{\rho} W')^{-1} S^{-1/2} \quad .$$

Pour la première de ces deux matrices, le fait que les corrélations soient toutes positives s'explique par le signe positif de α_1 dans $r_{ij} = \tanh(\alpha_1 d_{ij}^{-\lambda})$ $\forall i \neq j$. Alors, pour obtenir des corrélations négatives dans notre modèle, il faudra ajouter des variables dans la fonction r_{ij}.

Pour le modèle SAR(1), où $\varepsilon = \rho W \varepsilon + u$, la matrice des corrélations spatiales aura des termes tous positifs, lorsque les covariances seront toutes positives. Comme $E(\varepsilon\varepsilon') = \sigma^2 (I - \rho W)^{-1} (I - \rho W)^{-1'}$, tous les termes seront positifs lorsque l'inverse de $(I - \rho W)$ aura des termes tous positifs. Lorsqu'on se limite à la solution SAR(1) stable, l'expression stable $(I - \rho W)^{-1} = I + \rho W + \rho^2 W^2 + \ldots$ permet d'affirmer que pour $\rho > 0$, la matrice des corrélations spatiales aura des termes tous positifs. Ce résultat est d'ailleurs valable lorsque W est remplacé par $\underset{\sim}{W}$ ou M.

Remarque

Par souci de comparaison avec des résultats d'autres études dont l'estimation utilise le principe du MCG, on a présenté à l'équation (2) une mesure R^2 calculée selon la formule standard appliquée au modèle transformé $Py = PX\beta + P\varepsilon$ estimé par les MCO. Comme la colonne 1 de la matrice PX ne contient généralement pas de termes constants égaux à 1, le R^2 appliqué à cette équation est alors compris dans l'intervalle $-\infty \leq R^2 \leq 1$ et perd son interprétation de % de la variabilité de la variable transformée Py expliquée par la régression. Voir Dhrymes (1978), section 3.2.2 ou Judge (1980), section 6.2.2.

Pour un R^2 compris entre 0 et 1 et relié de façon biunivoque au test F de la régression dans son ensemble, consulter ces dernières références.

2) Autres spécifications

2.1) $z'_{ij}\ \alpha = d_{ij}^{-\lambda}\ (\alpha_1 + \alpha_2\ V^1_{ij})$

2.1.1) $V^1_{ij} = 1$, si $Pop_i > \overline{Pop}$ et $Pop_j > \overline{Pop}$; $= 0$, sinon

On l'a déjà dit, cette formulation est intéressante car elle permet de modéliser les corrélations des paires de zones à forte densité de population, d'une façon différente de celle des autres paires de zones. On va tout d'abord analyser les résultats de l'estimation.

(4)

$$y = \underset{(1,1091)}{8,2815} + \underset{(2,2907)}{0,0019}\ X$$

$$\hat{\alpha}_1 = \underset{(6,0966)}{2,5504}$$

$$\hat{\alpha}_2 = \underset{(-0,6572)}{-\ 0,0691}$$

$$\hat{\lambda} = \underset{(2,4417)}{0,5182)}$$

$R^2 = 0,2366$ $\quad \hat{\sigma}^2 = 53,5742$ $\quad \ln \mathcal{L} = -24,3016$ $\quad |R| = 1,4297\ E\text{-}24$.

Comme le coefficient $\hat{\alpha}_2$ n'est pas élevé, l'estimation à l'équation (4) ressemble beaucoup à celle présentée à l'équation (2). En fait, une importante population dans les deux zones ne semble par créer de corrélation spatiale dans les résidus. La vraisemblance n'a d'ailleurs pas assez augmenté pour

considérer comme différentes les spécifications $\alpha_1 d_{ij}^{-\lambda}$ et $d_{ij}^{-\lambda} (\alpha_1 + \alpha_2 V_{ij}^1)$. Il y a eu 46 évaluations de fonction et 12 itérations. La valeur obtenue par $\hat{\alpha}_2$ est trop petite pour permettre des corrélations négatives.

Finalement, bien que les résultats de cette estimation semblent avoir le même intérêt que ceux de l'estimation (2), ils n'ont pas été obtenus aussi facilement. En effet, pour les trouver, il a fallu donner les valeurs de départ suivantes : $\alpha_1^o = 0$, $\alpha_2^o = 0$, $\lambda^o = 1$ [1]. Pour des valeurs différentes de λ, l'algorithme se mit souvent à diverger de la façon présentée au tableau 3. A ce tableau suivant, on décrit le comportement de l'algorithme pour $\alpha_1^o = 0$, $\alpha_2^o = 0$ et $\lambda^o = 2$. Le phénomène qui en ressort représente bien le genre de problèmes qu'on a souvent rencontrés pour d'autres spécifications de V_{ij}^1.

Rappelons $\ln \mathcal{L} = -\frac{N}{2} \ln \sigma^{*2} - \frac{1}{2} \ln |R|$ [chapitre XI, équation (12)]. Dans le passage de l'évaluation 40 à l'évaluation 191, $\ln \mathcal{L}$ est passé de -24,01 à -12,81 pour des valeurs de paramètres relativement semblables. Si on décompose ce changement, on voit que pour $\Delta \sigma^{*2} = 0{,}822174$, $-\frac{N}{2} \ln \sigma^{*2}$ est passé de -36,735671 à -37,354111 alors que pour $\Delta |R| = 8{,}875253$ E-12 (infiniment petit), $-\frac{1}{2} \ln |R|$ est passé de 12,723877 à 24,544514. C'est donc la presque singularité de la matrice R plus que le changement dans les valeurs des paramètres qui a fait bouger la fonction de vraisemblance[2].

---

[1]. On donne des valeurs de départ zéro à α_1 et α_2 pour éviter de commencer les itérations avec un vecteur θ^o non permis. On suggérera dans ce qui suit d'utiliser la solution de l'estimation qui précède l'ajout d'une variable.

[2]. Le problème de la presque singularité de R serait probablement évité si on pouvait introduire une variable V_{ij}^1 avec un coefficient $\alpha_2 < 0$ assez élevé ce qui permettrait des corrélations spatiales résiduelles négatives.

TABLEAU 3

	Appel no	Fonction de vraisemblance	Valeur du déterminant	σ^{*2}	Valeur des paramètres		
					α_1	α_2	λ
1	18	- 25,752575	1,173992 E-5	11,218931	2,931237	0,081041	1,927366
2	31	- 24,413186	3,105129 E-11	16,585502	1,829870	1,769794	0,873543
3	40	- 24,011794	8,875253 E-12	16,874791	1,839320	1,950370	0,885227
4	49	- 22,966009	3,594610 E-13	17,614097	1,817308	1,997204	0,864301
5	113	- 17,318198	4,016014 E-18	17,686361	1,8155713861	2,0005735761	0,8626448027
6	151	- 14,892146	3,137233 E-20	17,686371	1,8155713659	2,0005736151	0,8626447834
7	191	- 12,808968	4,790264 E-22	17,696965	1,8155713658	2,0005736154	0,8626447833

Quel que soit le modèle étudié, pour un nombre limité de variables (1 ou 2) dans la forme linéaire $V'_{ij}\ \alpha$, le phénomène causé par la quasi-singularité de R risque très souvent de se produire. En effet, on l'a vu, la forme $\alpha_1\ d_{ij}^{-\lambda}$ donne aux corrélations un même signe lequel dépend du signe de α_1. Afin d'éviter de perdre un temps énorme à trouver des valeurs de départ admissibles qui en plus permettent à l'algorithme de converger, on propose maintenant une méthode qui, d'après nos expériences, semble assez bien fonctionner.

1) Estimer pour des valeurs de départ $\lambda^o = 1$ ou 2 et $\alpha_1^o = 0$ le modèle $\alpha_1\ d_{ij}^{-\lambda}$.

2) Pour le modèle $d_{ij}^{-\lambda}\ (\alpha_1 + \alpha_2\ V_{ij}^1)$, prendre comme valeurs de départ de λ et α_1 la solution de l'estimation contrainte ($\alpha_2 = 0$) précédente. Afin d'éviter la zone non permise, donner à α_2 la valeur de départ zéro.

3) A chaque fois qu'on ajoute une variable V_{ij}^h, on procède de la même façon qu'en 2).

Si on applique cette suggestion à notre problème, on obtient la solution présentée en (4) en 8 itérations et 22 appels de fonction dont 2 ont produit un vecteur θ non permis.

2.1.2) $v^1_{ij} = 1$, si $Pop_i > \overline{Pop}$ ou $Pop_j > \overline{Pop}$; = 0, sinon

Ici, on veut vérifier si le fait qu'au moins une des deux zones ait une forte densité de population peut créer de la corrélation spatiale dans les résidus. Pour des valeurs de départ $\alpha_1^o = 2,4725$ et $\lambda^o = 0,5612$, l'algorithme trouve la solution en 7 itérations et 20 évaluations de fonction dont 2 produisent un vecteur de paramètres non admissible. On obtient alors l'estimation suivante :

$$y = \underset{(0,8304)}{5,8042} + \underset{(2,6734)}{0,0023}\ X$$

(5)

$$\hat{\alpha}_1 = \underset{(6,2811)}{2,4448}$$

$$\hat{\alpha}_2 = \underset{(0,9473)}{0,0889}$$

$$\hat{\lambda} = \underset{(2,2733)}{0,5458}$$

$R^2 = 0,2804$ $\quad \hat{\sigma}^2 = 45,3257$ $\quad \ln \mathcal{L} = -24,1956$ $\quad |R| = 8,9355\ E\text{-}23$.

On arrive ici aux mêmes conclusions qu'à la section précédente. Pour notre modèle, seule la distance semble être un facteur explicatif de la corrélation spatiale résiduelle. Pour le vecteur de départ $[\alpha_1\ \alpha_2\ \lambda]^{o\prime} = [0\ \ 0\ \ 1]^{o\prime}$, l'algorithme a divergé alors que pour $[0\ \ 0\ \ 2]^{o\prime}$, il nous a conduit à un maximum local $\ln \mathcal{L} = -24,344889$ inférieur à celui obtenu en (5).

Remarque

Les estimations des sections 2.1.1) et 2.1.2) ont été refaites en remplaçant la population Pop par le revenu Rev. L'algorithme a alors été incapable de converger.

2.1.3) $V^1_{ij} = Pop_i \cdot Pop_j$

Le lien entre deux zones i et j dépend souvent de l'importance de la population et du revenu dans les deux zones. Afin de tester la signification de ces facteurs dans l'explication de la corrélation spatiale, on a proposé, au chapitre X, d'utiliser respectivement $V^1_{ij} = Pop_i \cdot Pop_j$ et $V^1_{ij} = Rev_i \cdot Rev_j$.

Ce genre de spécification nous a causé énormément d'ennuis. En fait, pour tous les vecteurs de départ essayés, l'algorithme a toujours divergé. Ainsi, pour le modèle qui nous intéresse, on est incapable de vérifier le bien-fondé de notre assertion.

2.1.4) $V^1_{ij} = 1$, si i et j sont contigües ; $= 0$ sinon

On aimerait ici tester si, en plus de la distance, la contiguïté est un facteur qui explique la corrélation spatiale résiduelle. Les complications de la section précédente se sont aussi produites ici. Toutefois, pour un λ fixé à 0,5612 ($\hat{\lambda}$ de la section 1)), on a été capable d'obtenir l'estimation suivante :

$$y = \underset{(1,1154)}{7,8258} + \underset{(2,3341)}{0,00198}\ X$$

$$\hat{\alpha}_1 = \underset{(4,9315)}{2,5428}$$

(6)

$$\hat{\alpha}_2 = \underset{(-0,1206)}{-0,0153}$$

$$\bar{\lambda} = 0,5612 \qquad \text{(- pour fixé à priori)}$$

$R^2 = 0,255$ $\quad \hat{\sigma}^2 = 47,4175$ $\quad \ln \mathcal{L} = -24,4245$ $\quad |R| = 4,369780\ E\text{-}23$.

Une fois la distance imposée dans la spécification, la contiguïté ne semble pas être un facteur qui explique le lien entre les erreurs du modèle. Ce résultat ne nous surprend toutefois pas. En effet, si on se reporte à l'estimation SAR(1), on se rappellera que lorsqu'on a imposé la contiguïté dans le schéma autorégressif, la distance est devenue non significative.

Il ressort de ces deux estimations, que dans le modèle qui nous intéresse, la distance et la contiguïté sont des facteurs fortement concurrents en ce sens qu'ils apportent une information équivalente sur la structure physique de l'étude. Ainsi lorsqu'on impose l'un, l'autre devient non significatif.

2.1.5) $$v_{ij}^1 = \frac{1}{|\ln Rev_i - \ln Rev_j|}$$

On l'a déjà remarqué, au cours de l'analyse des résultats du SAR(1), les zones qui forment la république d'Irlande semblent être regroupées spatialement selon leurs revenus. En effet, dans la région sud-est, on retrouve la majorité

des zones à haut revenu. Il serait alors intéressant de tester si la corrélation résiduelle entre les zones de revenus semblables est plus forte que celle de revenus différents, c'est-à-dire est-ce que la valeur absolue de la corrélation spatiale est une fonction directe de $V_{ij}^{1} = \frac{1}{|\ln Rev_i - \ln Rev_j|}$?

D'après l'évidence statistique des résultats présentés à l'équation (7) suivante, la similitude des revenus ne semble pas un phénomène significatif pour expliquer le lien entre les paires d'erreurs de l'étude.

$$y = \underset{(0,0306)}{0,1793} + \underset{(4,2534)}{0,0033}\ X$$

$$\hat{\alpha}_1 = \underset{(4,9866)}{2,1653}$$

(7)

$$\hat{\alpha}_2 = \underset{(1,0619)}{0,0009}$$

$$\hat{\lambda} = \underset{(1,9127)}{0,5233}$$

$R^2 = 0,4494 \qquad \hat{\sigma}^2 = 30,1207 \quad \ln \mathcal{L} = -23,4756 \qquad |R| = 8,713296\ E\text{-}18$

Pour des valeurs de départ qui correspondent à la solution de $\alpha_1 d_{ij}^{-\lambda}$ présentée à l'équation (2), on trouve la solution en 11 itérations et 36 appels dont aucun n'a produit un vecteur de solution θ non permis. L'examen de différentes grilles sur plusieurs intervalles de α_1 α_2 et λ semble confirmer la validité de ce résultat. L'estimation du modèle $V_{ij}^{1} = \frac{1}{|Rev_i - Rev_j|}$ conduit à des résultats très similaires.

3) Evaluation générale de nos estimations

Les estimations des spécifications étudiées à la section 2) n'ont pas permis de trouver des variables significatives autres que la distance. Ainsi, la forme la plus générale qu'on a estimé de façon fiable, pour la corrélation spatiale s'écrit :

$$(8) \qquad r_{ij} = \tanh\left(\alpha_1 \, d_{ij}^{-\lambda}\right) \quad .$$

A la section 1), on a pu constater que l'algorithme a trouvé une solution intéressante et ce en quelques itérations. Les grilles ont confirmé l'exactitude des résultats. Toutefois, cette forme (8) a l'inconvénient de produire des corrélations $\hat{r}_{ij}$ de même signe.

CONCLUSION

Afin de considérer le problème de corrélation spatiale dans les résidus de modèles de régression avec données spatiales (coupe transversale), nous avons suggéré deux approches. La première est basée sur une formulation de type autorégressif alors que la seconde est une approche de type MCG (moindres carrés généralisés) du maximum de vraisemblance où on s'intéresse à estimer les éléments de la matrice de corrélation des résidus.

Dans la première partie de la thèse, nous avons proposé d'estimer par une méthode du maximum de vraisemblance le vecteur joint des paramètres $\gamma' = [\beta' \ \sigma^2 \ \rho \ \theta']$ du modèle SAR(1) [autorégressif spatial d'ordre (1)] résiduel suivant :

$$y = X\beta + \varepsilon$$

$$\varepsilon = \rho W \varepsilon + u \quad ,$$

où ρ est un coefficient d'autocorrélation spatial et W une matrice de contiguïté où $w_{ij} = g(V^1_{ij},\ldots,V^P_{ij}, \theta_1,\ldots,\theta_H)$ est une forme fonctionnelle qui permet d'introduire dans le modèle les caractéristiques économico-géographiques de la région.

Au chapitre I, nous avons présenté les fondements théoriques et les origines de la formulation SAR(1). Au chapitre III, pour w_{ij} fixé à priori, nous avons fait ressortir les avantages théoriques et empiriques de deux types de normalisation sur la matrice W de contiguïté. Du point de vue théorique, pour $|\rho| < 1$, les matrices de contiguïté M et $\underset{\sim}{W}$ nous ont permis de carac-

tériser le modèle SAR(1) résiduel comme stable. Du point de vue empirique, les chapitres V à VII nous ont permis de constater que ce genre de normalisation pour $|\rho| < 1$, a contraint la recherche d'une solution $\hat{\gamma}$ qui maximise la vraisemblance de l'échantillon à une région pour laquelle le comportement de cette fonction est stable et uniforme.

Bien que l'on ait pu obtenir des estimations dont les paramètres sont significatifs, notre procédure n'a pu briser de façon satisfaisante la configuration spatiale résiduelle. Il faudrait peut-être passer à un ordre supérieur d'autorégression. On peut rapidement vérifer que dans ce cas, la normalisation est encore là, désirable.

La méthodologie développée dans cette première partie peut, comme vu aux chapitres VIII et IX, s'appliquer à des modèles d'explication de flux. Comme autre extension possible, on peut suggérer ici l'estimation des poids w_{ij} dans les formulations spatio-temporelles "spacetime" développées par Pfeifer et Deutsch (1980).

Dans la seconde partie (aux chapitres X à XII), nous avons développé une procédure du maximum de vraisemblance selon laquelle nous nous intéressions à modéliser les éléments de la matrice des corrélations des résidus. Après avoir corrigé, si nécessaire, l'hétéroscédasticité, le modèle homoscédastique transformé s'écrit $y^* = X^*\beta + \varepsilon^*$ avec $E(\varepsilon^*) = 0$ et $E(\varepsilon^*\varepsilon^{*\prime}) = \sigma^2 R$,
R étant la matrice $(N \times N)$ des corrélations entre les résidus de l'étude.
Afin de modéliser les corrélations spatiales r_{ij}, nous avons postulé la forme suivante : $r_{ij} = \tanh (z'_{ij}\, \alpha)$ avec $z'_{ij}\, \alpha = d_{ij}^{-\lambda} (\alpha_1 + \alpha_2\, v_{ij}^2 + \ldots + \alpha_H\, z_{ij}^H)$,
d_{ij} représentant la distance entre i et j.

Grâce à cette forme, nous limitons r_{ij} à l'intervalle unitaire $[-1 \leq r_{ij} \leq 1]$ et nous imposons la propriété désirable de corrélation nulle entre les erreurs de zones très éloignées. Le problème principal de notre approche réside dans le fait qu'on ne peut garantir pour R la propriété de matrice définie positive.

Les résultats d'estimation de cette procédure se sont avérés assez décevants, la spécification la plus générale s'étant limité à $z'_{ij}\ \alpha = \alpha_1\ d_{ij}^{-\lambda}$. Pour cette spécification, les estimations ont été assez satisfaisantes. Le principal mérite de notre démarche est probablement d'avoir ouvert la voie à d'autres recherches. Il faudrait diriger les efforts vers une méthodologie qui force R à être toujours définie positive.

APPENDICE I

DONNEES

Au tableau 1 suivant, on retrouve les observations pour les variables qui ont été utilisées lors de nos expériences empiriques.

TABLEAU 1 *

no lettre	noms	population[1]	% de la production autoconsommée[2]	ARA[3] en milles	valeur des ventes au détail[4] (£ 000)	revenu personnel total[5] (£ 000)	degré longitudinal[6]	degré latitudinal[7]
A	Carlow	33 342	8,6	3 664	2 962	7 185	6,8	52,7
B	Cavan	56 594	15,0	5 000	4 452	9 459	7,3	53,97
C	Clare	73 702	19,0	4 321	3 460	12 435	8,97	52,87
D	Cork	330 443	9,0	4 118	28 402	65 901	8,73	51,97
E	Donegal	113 842	27,0	7 500	7 478	17 626	7,92	54,9
F	Dublin	718 332	9,4	3 078	89 424	164 631	6,3	53,4
G	Galway	144 887	21,9	4 537	8 972	26 950	8,75	53,35
H	Kerry	116 458	17,0	5 140	6 341	20 510	9,58	52,13
I	Kildare	64 420	9,0	3 200	4 803	14 703	6,78	53,2
J	Kilkenny	61 668	8,0	3 708	4 321	13 585	7,22	52,55
K	Laoighis	45 069	10,3	3 455	3 128	9 280	7,32	53,0
L	Leitrim	33 470	23,1	5 000	1 885	5 709	8,0	54,22
M	Limerick	133 339	11,4	4 018	10 786	27 395	8,75	52,5
N	Longford	30 643	19,0	4 250	1 960	5 297	7,71	53,73
O	Louth	67 378	10,1	3 948	7 059	12 156	6,53	53,93
P	Mago	123 320	30,0	6 815	6 758	19 201	9,33	53,95
Q	Meath	65 122	8,7	4 008	3 356	14 512	6,67	53,63
R	Monaghan	47 088	13,0	4 500	3 960	8 396	6,93	54,15
S	Offaly	51 533	14,3	4 108	3 817	10 320	7,58	53,2
T	Roscommon	59 217	23,0	4 500	2 821	10 223	8,27	53,73
U	Sligo	53 561	22,0	5 997	3 535	9 461	8,67	54,17
V	Tipperary	123 822	9,0	3 926	9 226	26 424	7,88	52,62
W	Waterford	71 439	8,0	3 691	7 526	15 696	7,58	52,22
X	Westmeath	52 861	16,0	3 872	3 822	10 842	7,47	53,55
Y	Wexford	83 308	8,6	3 940	8 231	15 562	6,6	52,5
Z	Wicklow	58 473	10,2	3 600	4 865	11 921	6,35	52,98

Sources : [1] Census of Population of Ireland (Central Statistics Office, 1977), vol. 9, table 2.

[2] [3] [4] [5] Cliff et Ord (1981), table 8.1.

[6] [7] Geary (1968)

* Toutes les données fournies ici concernent l'année 1961

Les distances entre chacune des zones de l'étude ont été calculées à partir des coordonnées longitudinales (Y) et latitudinales (X) que l'on retrouve aux colonnes 6 et 7 du tableau I[1].

Etapes de construction des distances :

1. Variation (Δ) longitudinale et latitudinale exprimée en pouces. A partir de l'atlas consulté,

   1 degré latitudinal = 5,25"

   1 degré longitudinal = 3,00" .

   On calcule

   $$X_1 = \Delta \text{ lat}^{\circ} \times 5{,}25"$$

   $$Y_1 = \Delta \text{ long}^{\circ} \times 3{,}00" \ .$$

2. Conversion de pouces en milles :

   comme sur l'atlas, 3" équivalent à 40 milles, la distance en milles entre la zone i et j est calculée comme suit :

   $$d_{ij} = \frac{40}{3} \sqrt{X_1^2 + Y_1^2} \ ,$$

   où X_1 et Y_1, pour chaque paire i,j sont calculés en 1).

Suite à ces deux étapes, nous obtenons la matrice de distances suivante :

---

[1]. Les coordonnées de Dublin ont été approximées à partir d'un atlas. On a posé pour Dublin X = 53,4° et Y = 6,3°. Voir Geary (1968).

TABLEAU 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	0 0000													
B	91 1220	0 0000												
C	87 6119	101 9374	0 0000											
D	92 9800	151 2344	63 7272	0 0000										
E	160 2840	69 6638	148 1770	207 6434	0 0000									
F	52 9249	56 4979	113 0604	139 5272	123 3837	0 0000								
G	90 3009	72 4400	34 7333	96 6033	113 4658	98 0425	0 0000							
H	118 1417	157 8191	37 2591	35 7972	204 9541	138 4823	91 6264	0 0000						
I	35 0091	37 7741	90 3965	116 1775	127 4377	23 7622	79 4965	134 7368	0 0000					
J	19 8114	99 <515	73 4967	72 7772	164 8660	69 9606	82 9344	98 8772	48 7853	0 0000				
K	29 3574	67 9047	66 6244	91 3389	139 1481	49 4837	62 2261	108 9999	25 7402	31 7530	0 0000			
L	110 3828	29 9040	95 7167	153 3065	94 6937	84 6414	61 6864	152 9664	80 8010	114 2429	82 9843	0 0000		
M	79 2465	118 1203	27 3542	37 1086	171 2491	116 9032	59 5000	42 1076	92 7929	61 3000	67 0585	117 3012	0 0000	
N	80 7674	23 4776	78 5124	129 7801	82 3296	60 9473	49 3773	134 6813	52 5381	84 8936	53 4282	29 6623	95 6231	0 0000
O	86 7747 0 0000	30 9270	122 6026	162 9964	87 7597	38 2237	97 6412	173 3833	32 0493	100 4655	72 3641	60 2854	133 8112	49 2325
P	133 7822 112 0087	81 2121 0 0000	76 9592	140 6626	87 1964	127 1680	47 9817	127 7919	114 7181	129 3343	104 3380	94 5147	104 1177	66 6048
Q	65 3074 21 7938	34 6624 108 7323	106 2744 0 0000	142 4507	101 9961	21 8489	85 4775	156 7608	30 4199	78 7360	31 1938	63 2987	114 8002	42 1818
R	101 6 31 22 2072	19 4371 97 0155	121 1888 37 8566	168 7328 0 0000	65 7602	58 2348	91 8468	176 7200	66 7701	112 5991	81 9976	42 8315	136 3287	42 8696
S	46 8873 66 1454	33 0313 87 5000	60 2077 47 2331	97 6177 71 4020	119 7746 0 0000	53 0796	47 9434	109 9902	32 0000	47 7243	17 4402	66 5552	67 7507	37 4626
T	93 0368 70 9941	42 2810 45 1101	66 3931 64 3817	124 3664 81 1536	83 0880 46 2404	82 1161 0 0000	32 8035	123 6518	70 2038	92 6648	63 6805	29 3586	98 2148	22 4000
U	127 2142 87 2330	36 3601 30 3634	91 7878 88 4807	134 0187 69 6141	39 2555 80 6931	109 0316 34 7079	57 4891 0 0000	147 3662	101 6138	127 3717	98 1000	27 0276	116 9478	49 2240
V	43 3613 106 4185	97 3062 109 6887	66 9810 85 6799	96 8001 113 6416	139 6080 42 3363	93 5189 19 2306	61 8243 113 0080	76 1609 0 0000	39 8693	26 0309	34 7733	105 1097	35 7994	77 9970
W	45 8319 126 8546	123 0109 139 8757	71 8443 105 1781	49 2164 137 5791	188 0923 69 6000	97 1812 109 2440	91 9078 145 2941	80 2477 30 4631	75 6765 0 0000	27 2200	33 3817	134 0349	50 7383	105 8279
X	45 2571 42 0578	30 1761 79 4944	76 3882 32 4863	121 3423 47 2288	96 1990 24 8920	47 9634 34 3913	53 0796 64 7114	130 3982 67 1340	36 9034 93 2039	78 7107 0 0000	38 9647	45 1824	89 5751	15 8405
Y	16 1245 100 1592	106 6413 149 0869	98 2744 79 1495	92 9271 116 2518	176 1018 62 7306	64 1327 108 9745	104 2760 143 2551	121 9814 31 8945	49 3262 43 8269	29 0478 81 3221	43 3239 0 0000	126 4736	86 0000	96 8740
Z	26 6113 66 8884	79 0347 137 1825	105 0825 47 2662	118 9813 85 1226	148 3482 51 5554	29 4679 93 0295	99 4324 124 7026	142 2424 66 1052	33 0868 72 4430	46 0114 59 9921	38 8252 23 0565	103 3369 0 0000	101 7102	75 6017

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	O	P	Q	R	S	T	U	V	W	X	Y	Z		

APPENDICE II

RÉSULTATS CONCERNANT LES STATISTIQUES D'AUTOCORRÉLATION SPATIALE

1) Théorème Pitman-Koopmans

Soit $x = (x_1,\ldots,x_N)'$ un vecteur de variables aléatoires indépendantes et identiquement distribuées tel que $x \sim N(0, I)$, toute statistique $h(x)$ invariante dans le changement d'unité (scale free) où $h(\lambda x) = h(x)$ est distribuée indépendamment de $Q = \sum_{i=1}^{N} x_1^2 = x'x$.

Afin de prouver ce théorème, Cliff et Ord (1981) et Brandsma et Ketellapper (1979b) utilisent le fait que la fonction génératrice conjointe des moments (f.g.m.) de h et Q :

$$(1) \qquad M_{h,Q}(t_1, t_2) = E[e^{t_1 h + t_2 Q}] \quad ,$$

peut s'écrire comme le produit de leur (f.g.m.) marginales respectives :

$$(2) \qquad M_{h,Q}(t_1, t_2) = M_{h,Q}(t_1, 0) \cdot M_{h,Q}(0, t_2) \quad .$$

Plutôt que de prendre cette approche, on prouvera l'indépendance entre h(x) et Q en appliquant un théorème très général qui concerne l'indépendance entre une statistique exhaustive totale et une statistique libre[1].

Avant de passer à l'établissement de la preuve, rappelons certains concepts. Pour simplifier l'argumentation, on considérera l'estimation d'un seul paramètre θ par une statistique quelconque.

---

[1]. Voir Monfort (1982), chapitre V et plus particulièrement le théorème 4 de ce chapitre.

1.1) Statistique exhaustive totale et statistique libre

1.1.1) Statistique exhaustive totale

Soit $x_1,\ldots,x_N$, un échantillon aléatoire de taille N dont x_i possède la densité $f(\cdot,\theta)$. Une statistique $T = t(x_1,\ldots,x_N)$ est une variable aléatoire qui condense les N variables aléatoires $x_1,\ldots,x_N$ en une seule. Une statistique est exhaustive lorsque cette condensation ne produit aucune perte d'information concernant le paramètre à estimer θ. Une définition qui nous sera utile est :

Définition 1 : Statistique exhaustive (sufficient statistic)

Soit $x_1,\ldots,x_N$, un échantillon aléatoire tiré d'une population avec densité $f(\cdot,\theta)$. Une statistique $S = s(x_1,\ldots,x_N)$ est exhaustive si et seulement si la loi de distribution conditionnelle de toute statistique T étant donné S = s ne dépend pas de θ.

D'après cette dernière définition, pour prouver qu'une statistique T' n'est pas exhaustive, il s'agit de trouver une statistique T dont la loi de distribution conditionnelle $f_{T/T'}(\cdot)$ dépend de θ [1].

---

[1]. Une définition plus intuitive mais équivalente stipule qu'une statistique S est exhaustive si et seulement si la loi de distribution conditionnelle de $x_1,\ldots,x_N$ étant donné S = s ne dépend pas de θ, c'est-à-dire que toute l'information concernant θ est en S. De connaître S rend à cet égard la connaissance de l'échantillon inutile.

Remarque

Grâce au théorème de la factorisation[1] $f_{x_1,\ldots,x_N}(\cdot,\theta) = g(s,\theta) \cdot h(x_1,\ldots,x_N)$, il est possible de montrer qu'un estimateur du maximum de vraisemblance ne dépend de l'échantillon que par l'entremise de statistiques exhaustives. Le fait que toute transformation monotone d'une statistique exhaustive produit une statistique exhaustive pousse à conclure qu'un estimateur du maximum de vraisemblance est une statistique exhaustive.

Dans la recherche d'un estimateur non biaisé à variance minimale (MVUE), le théorème Rao-Blackwell établit qu'un estimateur T' non biaisé de θ, lequel est une fonction d'un certain nombre des statistiques exhaustives, possède une variance plus petite que tout autre estimateur T non fonction de statistiques exhaustives. Ce résultat est toutefois intermédiaire. Il ne dit pas que T' est MVUE, il dit simplement que dans la recherche d'un estimateur MVUE, il est préférable de se limiter aux fonctions des statistiques exhaustives.

Si on ajoute le concept de totalité à la statistique exhaustive S, le théorème Lehmann-Scheffé garantit que pour l'échantillon $x_1,\ldots,x_N$ aléatoire avec densité $f(\cdot,\theta)$, si S est exhaustive et totale, et si $T^* = t^*(S)$ est non biaisé d'une fonction $\tau(\theta)$, alors T^* est MVUE[2].

---

1. Voir Mood, Graybill et Boes (1974), p. 307.

2. Voir Mood, Graybill et Boes (1974), section 5.2 du chapitre VII. Il faut noter que cette propriété n'est pas toujours vraie. Une condition suffisante consiste à considérer la classe des densités exponentielles.

Il faut maintenant définir le concept de modèle total et de statistique totale.

Définition 2. Modèle total et statistique totale

Soit $x_1,\ldots,x_N$, un échantillon aléatoire avec densité $f(\cdot,\theta)$ où $\theta \in \Theta$, et soit $T = t(x_1,\ldots,x_N)$ une statistique. Le modèle est total si :

(3) $\quad E_\theta [z(T)] = 0 , \quad \forall \theta \in \Theta ,$

implique que $P[z(T) = 0] \equiv 1 \quad \forall \theta \in \Theta$ où $z(T)$ est une statistique. De plus, une statistique T est totale si le modèle est total.

Remarques

1. On note $E_\theta [z(T)]$ pour dire que l'espérance est prise sur une variable aléatoire dont la densité dépend de θ.

2. $E_\theta [z(T)] = \int_{x_1} \ldots \int_{x_N} z(t(x_1,\ldots,x_N))\, f(x_1,\theta),\ldots,f(x_N,\theta)\, dx_1,\ldots,dx_N,$

ou, de façon équivalente : $E_\theta [z(T)] = \int z(t)\, f_T(t,\theta)\, dt$.

D'après la définition 2, une statistique T est totale si la seule estimation de 0 pour $z(T)$ est $z(T) = 0$ avec probabilité 1.

A l'opposé de la statistique totale, il existe la statistique libre.

1.1.2 *Statistique libre*

Définition 3. Statistique libre

Une statistique T est libre lorsque sa fonction de densité ne dépend pas de θ. Autrement dit, une statistique est libre lorsqu'elle n'apporte pas d'information sur le paramètre θ d'intérêt. Nous sommes maintenant prêts à établir la preuve du théorème Pitman-Koopmans.

1.2) Démonstration du théorème Pitman-Koopmans

On veut décomposer notre preuve en deux étapes. Dans la première, par un théorème, on énonce et prouve l'indépendance entre une statistique libre et une statistique exhaustive totale. Dans la seconde, on applique ce théorème au cas qui nous intéresse ici.

Théorème 1

Soit S et T, deux statistiques où $x_1,\ldots,x_N$ est un échantillon aléatoire avec densité $f(\cdot,\theta)$. Si S est exhaustive et totale et T est libre, elles sont indépendantes.

Preuve

Comme T est libre, pour toute fonction z mesurable ≥ 0, $E_\theta [z(T)] = E [z(T)] \quad \forall\theta$. De plus, comme on a toujours

$E_\theta [z(T)] = E_\theta \{E [z(T)|S]\}$[1] $\forall\theta$, où on note $E [z(T)|S]$ car, puisque S est exhaustive, $f_{T/S}(\cdot)$ ne dépend pas de θ. Ces deux résultats nous permettent d'écrire que :

$$(4) \qquad E_\theta \{E [z(T)|S] - E [z(T)]\} = 0 \quad .$$

Comme S est une statistique totale, par la définition 2, on peut écrire que :

$$(5) \qquad E [z(T)|S] - E [z(T)] = 0 \quad , \qquad \text{avec probabilité 1 .}$$

Cette expression n'étant valable que lorsqu'on a indépendance, on obtient que T et S sont des statistiques indépendantes. En effet, par (5) :

$$\int_t z(t)\, f_{T/S}(\cdot)\, dt = \int_t z(t)\, f_T(\cdot)\, dt \quad ,$$

et donc

$$\int_t z(t) [f_{T/S}(\cdot) - f_T(\cdot)]\, dt = 0 \quad ,$$

lorsque $f_{T/S}(\cdot) = f_T(\cdot)$.

---

[1]. $E_\theta [z(T)] = \int_s \int_t z(t)\, f_{T,S}(\cdot, \theta)\, dt\, ds$

$= \int_s \int_t z(t)\, f_{T/S}(\cdot) \cdot f_S(\cdot, \theta)\, dt\, ds$

$= \int_s f_S(\cdot, \theta) [\int_t z(t)\, f_{T/S}(\cdot)\, dt]\, ds$

C.Q.F.A.

Appliquons maintenant ce théorème à notre cas : soit

$\underset{(N\times 1)}{x} = [x_1,\ldots,x_N]'$ un échantillon aléatoire tel que $x \sim N(0, \sigma^2 I_N)$. Il est reconnu que dans le cas de la normale, $\sum x_i^2$ est une statistique exhaustive et totale[1]. De plus, la statistique h(x) est libre car $h(\sigma^2 x) = h(x)$ et n'apporte aucune information concernant σ^2. D'après le théorème, pour tout σ^2 la statistique h(x) est indépendante de $\sum x_i^2$, ce qui comprend aussi le cas $\sigma^2 = 1$.

C.Q.F.D.

2) Espérance mathématique d'une forme quadratique

Théorème 1 (*Balestra (1972), chapitre X, 216-222*)

Soit Y vecteur de variables aléatoires à N composantes avec $E(Y) = 0$ et $V(Y) = \sigma^2 V$ définie positive. Soit B une matrice symétrique d'ordre N. L'espérance mathématique E(Y' B Y) est donnée par l'expression.

$$E(Y' B Y) = \sigma^2 \text{ tr } (B V) \quad . \tag{6}$$

Variance d'une forme quadratique

Théorème 2

Soit Y, du théorème 1 mais aussi caractérisé par une normale. Alors, la variance de Y' B Y est égale à :

$$V(Y' B Y) = 2 \sigma^4 \text{ tr } (B V B V) \quad , \tag{7}$$

---

[1]. Voir Mood, Graybill et Boes, chapitre VII, exercice 22 et théorème 9.

d'où on tire par $V(x) = E(x^2) - E^2(x)$:

(8) $$E\left[(Y' B Y)^2\right] = 2\sigma^4 \, tr\,(B V B V) + \sigma^4 \, tr^2\,(B V) \quad .$$

3) <u>Dérivation des moments de distribution des statistiques d'autocorrélation spatiale sur les résidus</u>

On a le modèle suivant :

(9) $$y = X\beta + \varepsilon \ ,$$

d'où les résidus MCO s'écrivent :

(10) $$e = \left[I - X\,(X'X)^{-1}\,X'\right] y = My = M\varepsilon \ ,$$

où $M'M = MM' = M$, d'où $E(ee') = \sigma^2 M$, $E(e'e) = E(\varepsilon' M\varepsilon) = \sigma^2 \, tr\, M = \sigma^2 (N-K)$.
Soit la mesure d'autocorrélation sur les résidus MCO :

(11) $$I_{MCO}^{CO} = \frac{N}{S_0} \frac{e' W e}{e'e} = \frac{N}{S_0} \frac{\varepsilon' M W M \varepsilon}{\varepsilon' M \varepsilon} \quad .$$

Les résultats établis à l'appendice II.1) permettent d'évaluer les moments de distribution de I_{MCO}^{CO} comme suit :

(12) $$E(I_{MCO}^{CO}) = \frac{N}{S_0} \frac{E(e' W e)}{E(e'e)} = \frac{N}{S_0} \cdot \frac{tr\,(W M)}{N-K} \quad ,$$

par l'application des théorèmes de l'appendice II.2). De la même façon, on a :

(13) $$E\{[\varepsilon' M\varepsilon]^2\} = \sigma^4 [2\,tr(M) + tr^2(M)]$$

$$\sigma^4 [2(N-K) + (N-K)^2]$$

$$\sigma^4 (N-K)(N-K+2) \quad ,$$

et $E\{[e'We]^2\} = \sigma^4 [2\,tr(WMWM) + tr^2(WM)]$, d'où

(14) $$E[(I^{CO}_{MCO})^2] = \frac{N^2 [2\,tr(WMWM) + tr^2(WM)]}{S_0^2 (N-K)(N-K+2)} \quad .$$

Dérivation pour la classe LUS ($\hat{e}$)

Dans cette classe, il faut délaisser K observations. Donc, par (11), on obtient :

(15) $$I^{CO}_{\hat{e}} = \frac{(N-K)}{S_0^*} \frac{\hat{e} W_0 \hat{e}}{\hat{e}'\hat{e}} \quad ,$$

où $S_0^* = \underset{[1\times(N-K)]}{i'} \underset{[(N-K)\times(N-K)]}{W_0} \underset{[(N-K)\times 1]}{i}$.

Grâce à Theil (1971), chapitre 5,

$$\underset{[(N-K)\times 1]}{\hat{e}} = \underset{[(N-K)\times N]}{C} \underset{[N\times 1]}{\varepsilon} \quad ,$$

où $\underset{(N\times 1)}{\varepsilon} \sim N(0, I)$.

$$CC' = I_{N-K} \quad \text{et} \quad \underset{(N\times N)}{C'C} = \underset{(N\times N)}{M} \quad ,$$

où $M = [I - X(X'X)^{-1}X']$.

Alors comme $E(\hat{e}'\hat{e}) = tr\, E(\hat{e}\hat{e}') = \sigma^2 (N-K)$

$$E[(\hat{e}'\hat{e})^2] = E[(\varepsilon'C'C\varepsilon)^2] = E[(\varepsilon'M\varepsilon)^2]$$

$$= \sigma^2 (N-K)(N-K+2) \qquad \text{par (13)} .$$

Donc, par les théorèmes de l'appendice II.2), on obtient :

$$(16) \qquad E(I_{\hat{e}}^{CO}) = \frac{N-K}{S_0^*} \frac{tr(W_0)}{(N-K)} = 0 ,$$

car $W_{ii} = 0 \quad \forall i$,

$$(17) \qquad E[(I_{\hat{e}}^{CO})^2] = \frac{(N-K)^2}{S_0^{*2}} \cdot \frac{2\, tr(W_0 W_0) + tr^2 W_0}{(N-K)(N-K+2)} = \frac{(N-K)}{(N-K+2)} \cdot \frac{2\, tr(W_0^2)}{S_0^{*2}} .$$

APPENDICE III

CONCEPTS UTILES LORS DE L'ÉTUDE DES PROPRIÉTÉS ASYMPTOTIQUES DES ESTIMATEURS

Les propriétés d'un estimateur peuvent être obtenues en examinant généralement les premiers moments de distribution de cet estimateur (moyenne, variance). Pour des échantillons de taille finie, les moments sont souvent difficiles à évaluer. Si ce n'est pas possible, alors on peut se contenter d'étudier les propriétés lorsque la taille de l'échantillon tend vers l'infini. C'est la théorie asymptotique.

1) Rappel sur certaines propriétés d'échantillon fini

Soit θ le paramètre de la population à étudier. L'estimateur $\hat{\theta}$ est *non biaisé* si $E(\hat{\theta}) = \theta$. La moyenne $E(\hat{\theta})$ de la loi de distribution de $\hat{\theta}$ permet de dire si $\hat{\theta}$ est centré sur θ, c'est-à-dire si on tire de la même population un nombre infini d'échantillons de même taille, est-ce que $\hat{\theta}$ nous donnera en moyenne la vraie valeur θ ($\sum_{i=1}^{N \to \infty} \frac{\hat{\theta}}{N} \approx \theta$) ?

Soit plusieurs estimateurs non biaisés de θ, l'économètre choisira logiquement l'estimateur dont la distribution est la plus concentrée autour de θ, c'est-à-dire l'estimateur qui a la plus petite variance. Un estimateur $\hat{\theta}$ non biaisé de θ est dit à variance minimale (MVUE) si sa variance est plus petite ou égale à la variance de tout autre estimateur non biaisé de θ, qu'il soit linéaire ou non.

Lorsqu'on connaît la loi de distribution de la population caractérisée par le paramètre θ, on peut établir un critère plus strict.

Proposition 1

Soit $X_1, X_2, \ldots, X_T$, un échantillon aléatoire tiré d'une population caractérisée par un paramètre θ. Si $\hat{\theta}$ est un estimateur non biaisé de θ, alors, sous des conditions générales :

$$(1) \qquad \text{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln \mathcal{L}(\theta | x_1, x_2, \ldots, x_T)}{\partial \theta^2}\right]} \quad .$$

L'expression de droite s'appelle la borne limite inférieure de Cramer et Rao (CRLB). Alors, d'après (1), dès que l'on a trouvé un estimateur non biaisé dont la variance est égale à cette borne, on sait alors qu'aucun autre estimateur non biaisé, tiré d'une même population, pourra avoir une variance plus petite.

La généralisation multivariée s'écrit : soit $X_1, X_2, \ldots, X_T$, un échantillon aléatoire tiré d'une population caractérisée par un vecteur de paramètre $\underset{(K\times 1)}{\gamma}$ si $E(\underset{(K\times 1)}{\hat{\gamma}}) = \underset{(K\times 1)}{\gamma}$, alors, sous des conditions générales :

soit $\sum_{\hat{\gamma}}$ la matrice variance-covariance de $\hat{\gamma}$, et soit :

$$I(\gamma) = -E\left[\frac{\partial^2 \ln \quad (\gamma | x_1, x_2, \ldots, x_T)}{\partial \gamma \partial \gamma'}\right] \quad , \text{ la matrice d'information, alors :}$$

$$(2) \qquad -I(\gamma)^{-1}$$

est une matrice semi-définie positive. Lorsque la matrice de variance-covariance de $\hat{\gamma}$ est égale à la borne de Cramer et Rao, $\hat{\gamma}$ est appelé un estimateur non biaisé efficace. La borne ne peut pas toujours être atteinte pour un estimateur non biaisé. L'estimateur à variance minimale est alors dit (MVUE).

2) Théorie asymptotique (voir Theil (1971), chapitre 8)

2.1) La limite en probabilité

Un estimateur $\hat{\theta}$ de θ sera convergent lorsque :

$$\text{(3)} \qquad \operatorname*{plim}_{N\to\infty} \hat{\theta}_N = \theta \quad ,$$

ce qui veut dire que si on augmente la taille de l'échantillon ($N\to\infty$), alors l'estimateur correspondant $\hat{\theta}_N$ aura une distribution qui dégénèrera vers le point fixe θ. C'est ce qu'on appelle la convergence.

Plus formellement, soit une suite de variables aléatoires $Y_1, Y_2, \ldots, Y_N$ .... avec fonctions de distribution $F_1(\cdot), F_2(\cdot), \ldots, F_N(\cdot)$ ... . Cette séquence converge en probabilité vers une constante c si :

$$\text{(4)} \qquad \lim_{N\to\infty} P\,[|Y_N - c| > \varepsilon] = 0 \quad , \qquad \forall \varepsilon > 0 \ ,$$

dans une notation plus compacte, (4) se réécrit :

$$\text{(5)} \qquad \operatorname*{plim}_{N\to\infty} Y_N = c \quad .$$

Pour un estimateur $\hat{\theta}$ de θ, ce qu'on étudie en théorie asymptotique, c'est le comportement de la séquence de variables aléatoires $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_N$ .... où $\hat{\theta}_N$ est l'estimateur $\hat{\theta}$ de θ pour un échantillon de taille N. Si $\operatorname*{plim}_{N\to\infty} \hat{\theta}_N = \theta$, alors $\hat{\theta}_N$ est un estimateur *convergent* de θ.

Basée sur l'inégalité de Chebyshev, une condition suffisante mais non nécessaire pour que $\operatorname{plim}_{N\to\infty} \hat{\theta}_N = \theta$ est :

$$\lim_{N\to\infty} E(\hat{\theta}_N) = \theta \quad , \tag{6}$$

et

$$\lim_{N\to\infty} \operatorname{var}(\hat{\theta}_N) = 0 \quad . \tag{7}$$

Ainsi, pour qu'un estimateur soit convergent, il suffit, lorsque $N\to\infty$, que le biais $E(\hat{\theta}) - \theta$ et la variance disparaissent. Judge et al. (1982) établissent, aux pages 266 et 267, différentes propriétés de l'opérateur plim pour des séquences de vecteurs, de matrices et de variables aléatoires, dont on reproduit les principales ici.

Soit α_N et β_N des séquences de variables aléatoires, α et β des constantes et soit $\operatorname{plim}_{N\to\infty} \alpha_N = \alpha$ et $\operatorname{plim}_{N\to\infty} \beta_N = \beta$ alors :

Règles

1. $\operatorname{plim}_{N\to\infty} \alpha_N \pm \beta_N = \operatorname{plim}_{N\to\infty} \alpha_N \pm \operatorname{plim}_{N\to\infty} \beta_N = \alpha \pm \beta$.

2. $\operatorname{plim}_{N\to\infty} \alpha_N \cdot \beta_N = \operatorname{plim}_{N\to\infty} \alpha_N \cdot \operatorname{plim}_{N\to\infty} \beta_N = \alpha \cdot \beta$.

3. Si $\beta_N \neq 0$ et $\beta \neq 0$, alors $\operatorname{plim}_{N\to\infty} \left(\frac{\alpha_N}{\beta_N}\right) = \frac{\operatorname{plim}_{N\to\infty} \alpha_N}{\operatorname{plim}_{N\to\infty} \beta_N} = \frac{\alpha}{\beta}$.

4. Si $\alpha_N \geq 0$ et $\alpha \geq 0$, alors $\operatorname{plim}_{N\to\infty} \sqrt{\alpha_N} = \sqrt{\operatorname{plim}_{N\to\infty} \alpha_N} = \sqrt{\alpha}$.

5. $\underset{N\to\infty}{\text{plim}}\ \alpha = \alpha$.

6. Si $\phi(\alpha_N)$ est une fonction continue, alors $\underset{N\to\infty}{\text{plim}}\ \phi(\alpha_N) = \phi(\alpha)$ (théorème de Slutsky).

Remarque

Ces concepts peuvent être étendus directement à la convergence de séquence vectorielle ou matricielle de variables aléatoires.

On a : soit la séquence de matrices A_N $(A_1, A_2, \ldots, A_N \ldots)$. Cette séquence converge vers une matrice $\underset{(R\times S)}{A}$, lorsque :

$$(8) \qquad \underset{N\to\infty}{\text{plim}}\ a_{ij,N} = a_{ij} \quad , \qquad \forall i,j \quad \begin{matrix} i = 1,\ldots,R \\ j = 1,\ldots,S \end{matrix} \ ,$$

ce que l'on note :

$$(9) \qquad \underset{N\to\infty}{\text{plim}}\ A_N = A \quad .$$

L'application multivariée des règles 1 à 6 est évidente, elle ne sera donc pas présentée ici. Il faut remarquer que la dimension de A_N ne dépend pas de N.

3) <u>Loi de distribution limite</u>

Tout comme dans le cas d'échantillon fini, afin de pouvoir discriminer entre différents estimateurs, on comparera ici les moments de leur loi de distribution limite. Theil (1971), section 8.2, présente comment on peut obtenir la loi de distribution limite d'un estimateur.

Revenons à notre séquence de variables aléatoires $Y_1, Y_2, \ldots, Y_N \ldots$ avec fonctions de distribution $F_1(\cdot), F_2(\cdot), \ldots, F_N(\cdot) \ldots$

3.1) Convergence en distribution

La séquence de variable aléatoire $Y_1, Y_2, \ldots, Y_N \ldots$ avec fonctions de distribution $F_1(\cdot), F_2(\cdot), \ldots, F_N(\cdot) \ldots$, converge en distribution vers une variable aléatoire Y avec loi de distribution F( ) si $F_N(\ )$ converge vers F( ) lorsque $N \to \infty$ pour tout point de continuité de F( ). F( ) est alors appelé la loi de distribution limite de cette séquence de variables aléatoires. On dit alors : $Y_N \xrightarrow{D} Y$.

Afin d'obtenir la distribution limite de la séquence de distribution $F_1(\ ), F_2(\ ), \ldots$, Theil (1971) se sert du théorème suivant[1] :

Theorème 1

Soit $\phi_1(t), \phi_2(t), \ldots$ la séquence de fonctions caractéristiques de $Y_1, Y_2 \ldots$. Alors $Y_1, Y_2 \ldots$ avec lois $F_1, F_2, \ldots$ converge en loi vers la loi limite F [caractérisée par $\phi(t)$] si, et seulement si, $\phi_N(t)$ converge vers $\phi(t)$ lorsque $N \to \infty$, $\forall t$, pourvu que $\phi(t)$ soit continu en $t = 0$.

---

1. $\phi(t) = E(e^{itx})$ où $i = \sqrt{-1}$ est la fonction caractéristique d'une variable aléatoire x. L'intérêt de la fonction caractéristique provient du fait qu'elle est toujours définie pour une variable aléatoire possédant une densité. De plus, il y a une correspondance biunivoque entre les fonctions de densité et les caractéristiques. Donc, si on connaît la fonction caractéristique d'une variable aléatoire x, on peut, en principe, déterminer sa fonction de densité. (Dhrymes, (1974), p. 13. Voir aussi Theil (1971), p. 78-79).

C'est-à-dire si les conditions du théorème sont respectées et si on reconnaît la loi de distribution qui correspond à la fonction caractéristique $\phi(t)$, cette loi de distribution est en fait la loi de distribution limite de Y_1, Y_2 ... Comme déjà dit, on écrira alors $Y_N \xrightarrow{D} Y$.

Avec l'aide des propositions suivantes sur les distributions limites, la tâche de dérivations de distributions limites devient moins complexe.

3.2) Propositions concernant l'établissement de la loi de distribution limite (Theil (1971), p. 370-71)

Règles

7. Soit (X_N, Y_N) $N = 1,2,\ldots$ une séquence de paires de variables aléatoires. Si $\operatorname{plim}_{N\to\infty} (X_N - Y_N) = 0$ et si Y_N a une distribution limite, alors X_N a la même loi de distribution limite.

8. Soit la même séquence de paires de variables aléatoires qu'en (7). Supposons que Y_N converge en distribution ($\xrightarrow{D}$) vers une variable aléatoire Y et X_N converge en probabilité ($\xrightarrow{P}$) vers une constante c. Alors : $X_N + Y_N \xrightarrow{D} c + Y$; $X_N \cdot Y_N \xrightarrow{D} cY$; $Y_N/X_N \xrightarrow{D} Y/c$, si $c \neq 0$.

9. Si $g(\ )$ est une fonction continue et si $X_N \xrightarrow{D} X$ [1], alors $g(X_N) \xrightarrow{D} g(X)$.

10. Si $g(\ )$ est une fonction continue, si $X_N - Y_N$ converge en probabilité vers zéro et $X_N \xrightarrow{D} X$, alors $g(X_N) - g(Y_N) \xrightarrow{P} 0$.

Note : ces règles sont aussi valables pour des séquences de vecteurs et matrices de variables aléatoires.

---

1. D'après la section 3.1) précédente, il est évident qu'ici X est une variable aléatoire possédant une loi de distribution limite. $X_N \xrightarrow{D} X$ veut dire : soit X_N possédant une distribution limite. Evidemment, lorsque X est une constante, la distribution limite est alors un point de masse avec probabilité 1.

4) Application dans le cas de l'estimation MCO (Theil (1971), chapitre 8)

Soit :

$$(10) \qquad \underset{(N\times 1)}{Y} = \underset{(N\times K)}{X} \; \underset{(K\times 1)}{\beta} + \underset{(N\times 1)}{u}$$

X est une matrice non stochastique où $E(u) = 0 \quad E(uu') = \sigma^2 I_N$ et $\lim_{N\to\infty} \frac{1}{N} X'X = Q$ (matrice définie positive), soit $b = (X'X)^{-1} X'Y$ et $e = Y - Xb$. Alors :

i) b est convergent

preuve : comme $E(b) = \beta$ et $V(b) = \frac{\sigma^2}{N} \left(\frac{X'X}{N}\right)^{-1}$,

$$\lim_{N\to\infty} E(b) = \beta \ ,$$

$$\lim_{N\to\infty} V(b) = \lim \frac{\sigma^2}{N} \cdot Q^{-1} \ .$$

Comme $\frac{\sigma^2}{N}$ converge vers zéro, b est convergent par les équations (6) et (7).

ii) b a une distribution asymptotique normale

Comme $b = \beta + (X'X)^{-1} X' u$, alors :

$$(11) \qquad \sqrt{N}\,(b - \beta) = \left(\frac{X'X}{N}\right)^{-1} \cdot \frac{X'u}{\sqrt{N}} \ .$$

$\sqrt{N}\,(b - \beta)$ est donc le produit d'une matrice dont la limite lorsque $N\to\infty$ est Q^{-1}, par un vecteur aléatoire $\frac{X'u}{\sqrt{N}}$, dont on considère être capable de trouver la loi de distribution limite, ce que l'on note :

(12) $$\frac{X'u}{\sqrt{N}} \xrightarrow{D} z \quad .$$

D'après la règle 8, $\sqrt{N}\,(b-\beta)$ convergera en distribution vers la variable aléatoire $Q^{-1}z$, c'est-à-dire :

(13) $$\sqrt{N}\,(b-\beta) \xrightarrow{D} Q^{-1}z \quad .$$

Par l'application du théorème 1, Theil (1971), p. 380-81, montre que sous des conditions supplémentaires, on aura : $z \sim N(0, \sigma^2 Q)$. Alors, par (12) et (13) :

(14) $$\sqrt{N}\,(b-\beta) \sim N(0, \sigma^2 Q^{-1}) \quad .$$

iii) e_t converge en distribution vers u_t $\qquad t = 1,\ldots,N$

(15) $$e = Y - Xb = MY = Mu \quad ,$$

car $MX = 0$, où $M = I_N - X(X'X)^{-1}X'$. Alors :

$$E(e-u) = -X(X'X)^{-1}X'\,E(u) = 0$$

(16)
$$V(e-u) = \frac{\sigma^2}{N} X \left(\frac{X'X}{N}\right)^{-1} X' \quad .$$

Comme e et u sont des vecteurs dont la longueur dépend de N, alors il n'est pas possible d'évaluer $\underset{N\to\infty}{\text{plim}}\,(e-u) = 0$. Par contre, on peut montrer que chaque erreur $e_t \xrightarrow{D} u_t$. En effet, $E(e_t - u_t) = 0$ par (16) et

$$V(e_t - u_t) = \frac{\sigma^2}{N} [x_{t1},\ldots,x_{tK}] \left(\frac{X'X}{N}\right)^{-1} \begin{bmatrix} x_{t1} \\ \vdots \\ x_{tK} \end{bmatrix}, \text{ donc}$$

$$\lim_{N\to\infty} V(e_t - u_t) = \lim_{N\to\infty} \frac{\sigma^2}{N} \cdot x_{t\cdot} Q^{-1} x'_{t\cdot} = 0 \cdot [c_t \neq 0],$$

donc, par les équations (6) et (7), $\operatorname{plim}_{N\to\infty} (e_t - u_t) = 0$, et donc par la règle 7, e_t converge en distribution vers u_t.

5) Efficacité asymptotique

Tout comme dans les échantillons de taille finie, afin de discriminer entre différents estimateurs, on va établir le concept d'efficacité asymptotique. Pour ce faire, on se limitera à la classe des estimateurs convergents et asymptotiquement normalement distribués. Dans cette classe, celui qui atteint la bonne limite de Cramer et Rao asymptotique, est asymptotiquement efficace. (Voir Dhrymes (1974), lemme 17, page 129.)

APPENDICE IV

PROPRIÉTÉS DES ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE

1) Echantillon de taille finie

a) Efficacité

S'il existe un estimateur non biaisé dont la variance atteint la borne inférieure de Cramer et Rao, il provient d'une méthode du maximum de vraisemblance.

Par exemple, si on pose la normalité des erreurs dans le modèle linéaire de l'appendice III, $b = (X'X)^{-1} X'Y$ est identique à l'estimateur du maximum de vraisemblance. b est donc efficace. On peut montrer qu'aucun estimateur non biaisé n'atteint la borne pour σ^2. Alors $\tilde{\sigma}^2 = \frac{e'e}{N-K}$ est MVUE.

b) Propriété d'invariance

Si $\tilde{\theta}$ est un estimateur du maximum de vraisemblance de θ, alors l'estimateur du maximum de vraisemblance d'une fonction $g(\theta)$ est $g(\tilde{\theta})$.

2) Echantillon asymptotique

Lorsque l'échantillon est aléatoire, sous certaines conditions de régularité[1], les estimateurs du maximum de vraisemblance sont convergents, ont une distribution asymptotique normale et sont asymptotiquement efficaces, c'est-à-dire $\underset{(K\times 1)}{\text{plim}}\ \tilde{\theta} = \underset{(K\times 1)}{\theta}$.

(1) $\tilde{\theta} \sim AN\ (\theta, P^{-1})$

où $R = -E\left[\frac{\partial^2 \ln \mathcal{L}}{\partial\theta\partial\theta'}\right]$ (voir Theil (1971), sections 8.4 et 8.5 ou Dhrymes (1974), section 3.6).

---

1. Ces conditions sont toutes respectées dans le cas de la normale (Theil (1971), p. 387).

Donc, quelle que soit la loi de distribution de la population, asymptotiquement, $\tilde{\theta} \sim N$ et la matrice de variance-covariance asymptotique de $\tilde{\theta}$ atteint la borne R^{-1}.

Ceci s'écrit encore :

$$(2) \qquad \sqrt{N}\,(\tilde{\theta} - \theta) \sim N(0,\ N\,R^{-1}(\theta))$$

où $N\,R^{-1}(\theta) = E\left[\frac{1}{N}\,\frac{\partial^2 \ln \mathcal{L}(\theta|\ldots)}{\partial\theta\partial\theta'}\right]^{-1}$.

De plus, par l'application du théorème de Khinchine[1], Dhrymes (1974, p. 118) obtient :

$$(3) \qquad \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\,\frac{\partial^2 \ln \mathcal{L}(\theta|\ldots)}{\partial\theta\partial\theta'}\right] = E\left[\frac{1}{N}\,\frac{\partial^2 \ln \mathcal{L}(\theta|\ldots)}{\partial\theta\partial\theta'}\right] \quad .$$

Comme on l'a souligné dans le texte, ces résultats ont été établis à partir d'un échantillon aléatoire. Pour ce même résultat dans le cas d'une loi de probabilité jointe d'un vecteur normal $\varepsilon \sim N(0,\ \sigma^2\,\Omega)$ [voir Magnus (1978)].

---

[1]. Soit $\{x_t \quad t = 1,2,\ldots\}$ une séquence (i.i.d.) avec moyenne finie μ.

$$\operatorname*{Plim}_{T\to\infty} \bar{x}_t = \operatorname{plim} \frac{1}{T} \sum_{t=1}^{T} x_t = \mu \ .$$

REMERCIEMENTS

Au terme de cette recherche, j'aimerais remercier tout particulièrement le professeur Marcel G. Dagenais qui, en tant que directeur de cette recherche, a fait montre d'une disponibilité exemplaire. Ses conseils se sont avérés extrêmement pertinents et utiles.

Mes remerciements vont aussi aux professeurs Marc J.I. Gaudry et Jean-Marie Dufour qui, en agissant comme lecteurs, ont fourni des critiques à la fois judicieuses et constructives.

La très grande attention qu'a porté le professeur Christian Gourieroux à la lecture de la version préliminaire m'a grandement touché et motivé. Ces précieux conseils m'ont aidé à améliorer les fondements théoriques de cette recherche.

Je remercie mon épouse France et mes deux enfants d'avoir eu la patience nécessaire pour supporter une vie sociale trop souvent perturbée au cours de ce long périple qu'est l'élaboration d'une thèse de doctorat. Leur soutien moral a été grandement apprécié.

Je tiens enfin à remercier Mme Lucette Boyer Psarianos pour son amitié et l'inébranlable détermination dont elle a fait preuve en dactylographiant maintes et maintes fois ce texte, sans jamais se décourager et ce en sachant garder patience et enthousiasme tout en me procurant support et encouragement.

BIBLIOGRAPHIE

Abramowitz, M. et I.A. Stegun, 1972, Handbook of Mathematical Functions, Dover, New York.

Anderson, T.W., 1971, The Statistical Analysis of Time Series, John Wiley and Sons.

Anselin, L., 1982, "A Note on Small Sample Properties of Estimators in a First Order Spatial Autoregressive Model", Environment and Planning, A, 14, 8, (August), 1023-30.

Balestra, P., 1972, Calcul matriciel pour économistes, Castella, Suisse.

Bartels, C.P.A., 1979, "Operational Statistical Methods for Analyzing Spatial Data", dans Exploratory and Explanatory Statistical Analysis of Spatial Data, édité par Bartels, C.P.A. et R.H. Ketellapper, 1979, Martinus Nijhoff Publishing.

Bartels, C.P.A. et L. Hordijk, 1977, "On the Power of the Generalized Moran Contiguity Coefficient in Testing for Spatial Autocorrelation Among Regression Disturbances", Regional Science and Urban Economics, 7, 83-101.

Bivand, R.S., 1984, "Regression Modeling with Spatial Dependance : an Application of Some Class Selection and Estimation Methods", Geographical Analysis, 16, 1, (January), 25-37.

Blommestein, H.J., 1983, "Specification and Estimation of Spatial Econometric Models. A Discussion of Alternative Strategies for Spatial Economic Modeling", Regional Science and Urban Economics, 13, 251-270.

Bodson, P. et D. Peeters, 1975, "Estimation of the Coefficents of a Linear Regression in the Presence of Spatial Autocorrelation. An Application to the Belgian Labour-Demand Function", Environment and Planning, A, 7, 455-472.

Bowden, R., 1973, "The Theory of Parametric Identification", Econometrica, 41, 6, (November), 1069-74.

Brandsma, A.S. et R.H. Ketellapper, 1979a, "A Biparametric Approach to Spatial Autocorrelation", Environment and Planning, A, 11, 51-58.

Brandsma, A.S. et R.H. Ketellapper, 1979b, "Further Evidence on Alternative Procedures for Testing of Spatial Autocorrelation Among Regression Disturbances", dans Exploratory and Explanatory Statistical Analysis of Spatial Data, édité par Bartels, C.P.A. et R.H. Ketellapper, 1979, Martinus Nijhoff Publishing.

Burguete, J.F., A.R. Gallant et G. Souza, 1982, "On Unification of the Asymptotic Theory of Nonlinear Econometric Models", Econometric Reviews, 1, 2, 151-190.

Burridge, P., 1980, "On the Cliff-Ord Test for Spatial Correlation", Journal of the Royal Statistical Society, Series B, 42, 1, 107-8.

Burridge, P., 1981, "Testing for a Common Factor in a Spatial Autoregressive Model", Environment and Planning, A, 13, 795-800.

Cliff, A.D. et J.K. Ord, 1969, "The Problem of Spatial Autocorrelation" dans London Papers in Regional Science, vol. 1, Studies in Regional Science, London, édité par Scott, A.J.

Cliff, A.D. et J.K. Ord, 1971, "Evaluating the Percentage Points of a Spatial Autocorrelation Coefficient", Geographical Analysis, 3, 51-62.

Cliff, A.D. et J.K. Ord, 1973, Spatial Autocorrelation, Pion, London.

Cliff, A.D. et J.K. Ord, 1975a, "Model Building and the Analysis of Spatial Pattern in Human Geography (with Discussion)", Journal of the Royal Statistical Society, Series B, 37, 3, 297-349.

Cliff, A.D. et J.K. Ord, 1975b, "The Choice of a Test for Spatial Autocorrelation", dans Display and Analysis of Spatial Data, NATO Advance Study Institute, édité par Davis, J.C., John Wiley and Sons.

Cliff, A.D. et J.K. Ord, 1975c, "Space-Time Modelling with an Application to Regional Forecasting", Transaction of the Institute of British Geographers, 64, (March), 119-128.

Cliff, A.D., et J.K. Ord, 1981, Spatial Processes, Models and Application, Pion, London.

Dacey, M.F., 1965, "A Review of Measures of Contiguity for Two and K-Color Maps", dans Spatial Analyses : A Reader in Statistical Geography, édité par Berry, B.J.L. et D.F. Marble, Prentice-Hall, Englewood Cliffs, 479-495.

Davidon, W.C., 1959, "Variable Metric Method for Minimisation", A.E.C. Research and Development Report, ANL-5990.

Dhrymes, P.J., 1974, Econometrics, Statistical Foundations and Applications, Springer-Verlag.

Dhrymes, P.J., 1978, Introductory Econometrics, Springer-Verlag.

Doreian, P., 1981, "On the Estimation of Linear Models with Spatially Distributed Data", Chapitre 11 dans Sociological Methodology, édité par Leinhart, S.S., The American Sociological Association, School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh.

Durbin, J., 1960, "Estimation of Parameters in Time-Series Regression Models", Journal of the Royal Statistical Society, Series B, 22, 1, 139-153.

Durbin, J. et G.S. Watson, 1950, "Testing for Serial Correlation in Least Squares Regression I", Biometrika, 37, 409-28.

Fisher, W.D., 1971, "Econometric Estimation with Spatial Dependence", Regional and Urban Economics, Operational Methods, 1, 1, 19-40.

Fletcher, R. et M.J.D. Powell, 1963, "A Rapidly Convergent Descent Method for Minimisation", Computer Journal, 6, 163-68.

Fuller, W.A., 1976, Introduction to Statistical Time Series, John Wiley and Sons.

Gallant, R.A. et A. Holly, 1980, "Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation", Econometrica, 48, 3, (April), 697-720.

Gatrell, A.C., 1977, "An Introduction to Spatial Autocorrelation and its Geographical Applications", Discussion Papers in Geography, number 2, (April), University of Salford, Angleterre.

Gatrell, A.C., 1979a, "Autocorrelation in Spaces", Environment and Planning, 11, 507-16.

Gatrell, A.C., 1979b, "The Autocorrelation Structure of Central-Place Populations in Southern Germany, dans Statistical Application in the Spatial Sciences, Chapitre 4, édité par Wrigley, N.

Geary, R.C., 1968, "The Contiguity Ratio and Statistical Mapping" dans Spatial Analyses : A Reader in Statistical Geography, édité par Berry, B.L.J. et D.F. Marble, Englewood Cliffs, N.J., Prentice-Hall.

Gebizlioglu, Ö.L., 1982, "Multidimensional Random Processes and Spatial ARMA Family Models", Proceedings of Statistics Seminars, Special Issues of METU Studies in Development, 19-54.

Gebizlioglu, Ö.L., 1983, "On the Modelling, Estimation and Hypothesis Testing for Spatial ARMA Processes", Special Topics, International Time Series Meeting in Toronto, Canada.

Gourieroux, C., A. Monfort et A. Trognon, 1984, "Pseudo Maximum Likelihood Methods : Theory", Econometrica, 52, 3, (May), 681-701.

Gourieroux, C. et A. Monfort, 1983, Cours de séries temporelles, Collection Economie et Statistiques avancées, Economica, Paris.

Granger, C.W.J., 1969, "Spatial Data and Time Series Analysis", dans London Papers in Regional Science, vol. 1, Studies in Regional Science, London, édité par Scott, A.J.

Granger, C.W.J. et P. Newhold, 1977, Forecasting Economic Time Series, Economic Theory and Mathematical Economics, Academic Press, New York.

Griffith, D., 1980, "Towards a Theory of Spatial Statistics", Geographical Analysis, 12, 4, 325-39.

Haining, R.P., 1977, "Model Specification in Stationary Random Fields", Geographical Analysis, IX, 107-29.

Haining, R.P., 1978a, "Estimating Spatial Interaction Models", Environment and Planning, A, 10, 305-20.

Haining, R.P., 1978b, "Specification and Estimation Problems in Model of Spatial Dependence", Northwestern University Studies in Geography, Evanston.

Haining, R.P., 1978c, "The Moving Average Model for Spatial Interaction", Transaction of the Institute of British Geographers, 3, 202-25.

Haining, R.P., 1978d, "A Spatial Model for High Plains Agriculture", Annals of the Association of American Geographers, 68, 4 (December) 493-504.

Haining, R.P., 1980, "Spatial Autocorrelation Problems", dans Geography and The Urban Environment, Progress in Research and Application, édité par Herbert, D.T. et R.J. Johnston, John Wiley and Sons, Chapitre 1.

Hanna, F.A., 1966, "Effects of Regional Differences in Taxes and Transportation Charges on Automobile Consumption", dans Papers on Regional Statistical Studies, édité par S. Ostry et T.K. Rhymes, Toronto University Press.

Harvey, A.C., 1976, "Estimating Regression Models with Multiplicative Heteroscedasticity", Econometrica, 44, 461-65.

Hepple, L., 1976, "A Maximum Likelihood Model for Econometric Estimation with Spatial Series", dans Theory and Practice in Regional Science, édité par Masser, I., Pion, 90-104.

Hepple, L.W., 1979, "Bayesian Analysis of the Linear Model with Spatial Dependance", dans Exploratory and Explanatory Statistical Analysis of Spatial Data, édité par Bartels, C.P.A. et R.H. Ketellapper, 1979, Martinus Nijhoff Publishing.

Hordijk, L., 1974, "Spatial Correlation in the Disturbances of a Linear Interregional Model", Regional and Urban Economics 4, 117-140.

Hordijk, L., 1979, "Use of Econometric Relations in Space, Papers of the Regional Science Association, 42, 100-15.

Hordijk, L. et J.H.P. Paelinck, 1976, "Some Principles and Results in Spatial Econometrics", Recherches économiques de Louvain, 42, 3, 175-97.

Jenrich, R.I., 1969, "Asymptotic Properties of Non-Linear Least Squares Estimators", The Annals of Mathematical Statistics, 40, 2, 633-43.

Johnston, J., 1984, Econometric Methods, 3rd Edition, McGraw-Hill Book Co.

Judge, G.G., W.E. Griffiths,, R.C. Hill et T.C. Lee, 1980, The Theory and Practice of Econometrics, John Wiley and Sons.

Judge, G.G., R.C. Hill, W.E. Griffiths, H.Lütkepohl et T.C. Lee, 1982, Introduction to the Theory and Practice of Econometrics, John Wiley and Sons.

Kendall,M.G. et A. Stuart, 1963, "The Advanced Theory of Statistics" vol. 1, 2, 3, Charles Griffin & Company, London.

Magnus, J.R., 1978, "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix", Journal of Econometrics 7, 281-312.

Malinvaud, E., 1978, Méthodes statistiques de l'économétrie, 3ième édition, Dunod.

Martin, R.L., 1974, "On Spatial Dependence, Bias and the Use of First Spatial Differences in Regression Analysis", AREA, 6, 3, 185-194.

Monfort, A., 1982, Cours de Statistique Mathématique, Economica.

Mood, A.M., F.A. Graybill et D.C. Boes, 1974, Introduction to the Theory of Statistics, 3ième édition

Moran, P.A.P., 1948, "The Interpretation of Statistical Maps", Journal of the Royal Statistical Society, Series B, 10, 243-51.

Moran, P.A.P., 1950, "A Test for Serial Independence of Residuals", Biometrika, 37, 178-181.

Oden, N.L., 1984, "Assessing the Significance of a Spatial Correlogram", Geographical Analysis, 16, 1, 1-16.

Ord, J.K., 1975, "Estimation Methods for Models of Spatial Interaction", Journal of the American Statistical Association 70, 349, 120-26.

Ord, J.K., 1981, "Towards a Theory of Spatial Statistics : A Comment", Geographical Analysis, 13, 1, 86-91.

O'Sullivan, P.M., 1968, "Accessibility and the Spatial Structure of the Irish Economy", Regional Studies, 2, 195-206.

Pfeifer, P.E. et S.J. Deutsch, 1980, "A Three-stage Iterative Procedure for Space Time Modeling", Technometrics 22, 35-47.

Phillips, G.D.A. et A.C. Harvey, 1974, "A Simple Test for Serial Correlation in Regression Analysis", Journal of the American Statistical Association, 69, 935-39.

Rothenberg, T.J., 1971, "Identification in Parametric Models", Econometrica, 39, 577-92.

Theil, H., 1971, Principles of Econometrics, John Wiley and Sons.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1, (January), 1-25.

White, H. and I. Domowitz, 1983, "Nonlinear Regression with Dependent Observations", Working paper, Department of Economics, University of California, San Diego.

Whittle, P., 1954, "On Stationary Processes in the Plane", Biometrika 41, 434-49.